信号与噪声

[美] 纳特·西尔弗（Nate Silver）◎著　胡晓姣　张　新　朱辰辰◎译

the signal and the noise

why so many predictions fail— but some don't



中信出版社·CHINACITICPRESS

信号与噪声

[美]纳特 · 西尔弗著
胡晓姣 张新 朱辰辰译
中信出版社

图书在版编目（CIP）数据
信号与噪声 /（美）西尔弗著；胡晓姣，张新，朱辰辰译 .—北京：中信出版社，2013.8
书名原文：The Signal and the Noise
ISBN 978–7–5086–4114–0
I. ①信… ;II. ①西… ②胡… ③张… ④朱… ;III. ①经济预测学 ;IV. ① F201
中国版本图书馆 CIP 数据核字（2013）第 155105 号

信号与噪声
著者：[美] 纳特 · 西尔弗
译者：胡晓姣; 张新; 朱辰辰 策划推广：中信出版社（China CITIC Press） 出版发行：中信出版集团股份有限公司
（北京市朝阳区惠新东街甲 4 号富盛大厦 2 座; 邮编 ;100029）
（CITIC Publishing Group）
承 印 者：

开本：787mm×1092mm; 1/16 印 张：26.75 字 数：500 千字 版 次：2013 年 8 月第 1 版;;
印 次：2013 年 8 月第 1 次印刷
京权图字：01–2013–1998 广告经营许可证：京朝工商广字第 8087 号
书 号：ISBN 978–7–5086–4114–0/F·2968
定 价：69.00 元

中信出版社官网：http://www.publish.citic.com/
官方微博：http://weibo.com/citicpub
更多好书，尽在中信飞书 App：http://m.feishu8.com（中信电子书直销平台）

目录

献给我的父母

引言

本书的内容涵盖了信息技术和科学进步，也包括了自由市场、商业竞争以及观念革新。本书罗列了许多事物，这些事物使得人类比计算机更聪明，书中同样列举了人类所犯的错误。本书还描述了我们如何一步一步地学习认识客观世界，也解释了为什么有时会出现历史倒退的情况。

这是一本有关预测学方面的书，这一学科是涉及上述所有内容的交叉学科，旨在研究为什么有些预测很准确，而另外一些预测却失败了。我真诚地希望这本书能使你在规划未来时更有远见，而避免目光短浅。

信息越多，问题越多

人类最初的信息技术革命并非始于微型芯片的发明，而是以印刷机的诞生为开端的。1440 年，约翰内斯·古腾堡发明了印刷机，这项发明使普通民众能方便地了解信息，由此产生的思想洪流带来了前所未有的结果和影响。印刷机的发明点燃了 1775 年的工业革命之火，也促成了人类文明的迅猛发展——由从前那种科学和经济几乎止步不前的状态迅速跨越到呈指数型增长的状态，还带来了我们今天所熟悉的变化。除此之外，这一发明促成了一些历史事件，这些事件开启了欧洲启蒙运动，也加速了美利坚合众国的建立。

但是，印刷机发明之初并没有引发上述各种巨变，倒是可能催生了另外的事件，如持续了几百年的“圣战”。当人类开始相信他们可以预知和选择自己的命运时，人类历史上最血腥的时代也就开始了。

早在古腾堡发明印刷机之前，书籍就已经存在了，但当时书的数量和读书的人都很少，书籍只是贵族阶层的奢侈品。因为抄写员每次只能抄写一份副本，复制一份原稿的费用大约是每 5 页 1 弗洛林（一种金币，1 弗洛林约合 200 美元），因此像你现在读的这本书，在当时可能要花费 20 000 美元才能得到。而且，经过无数次的抄写，书中难免有大量的抄写错误，这些错误代代相传，成倍增加，甚至还会演变成与原意相反的错误。

这就使得知识的积累变得极其困难。要想阻止以文字记录的知识的不断减少，就需要付出巨大的努力，因为书籍腐烂的速度远远大于其生成的速度。只有几个版本的《圣经》和少量的哲学论述——比如柏拉图和亚里士多德的文章——被保留了下来，其他那些数不清的智慧，由于缺乏记载动机，都遗失在那个年代了。

过去，也许人们觉得对于知识的追求即使不全是无用功，似乎也没什么实际价值。如今世事瞬息万变，我们总会有一种“世事无常”的感

觉，而对于我们的前辈来说，这种“无常的世事”则更受关注。《圣经·传道书》中有一句唯美的诗句：“日光之下无新事”，可事实却并非如此，之所以说“无新事”，并不是因为每件事都已经被发掘出来了，而是因为所有事情都将被遗忘。

印刷机永久而深刻地改变了这一状况。几乎是一夜之间，一本书的成本就骤降近 300 倍，书的售价从相当于今天 20 000 美元的价格剧降至 70 美元。印刷机迅速在欧洲普及，到 1470 年，印刷机已从德国传播到罗马、西班牙塞维利亚、法国巴黎和瑞士巴塞尔，随后的 10 年里，几乎所有的欧洲主要城市里都有了印刷机的影子。印刷机问世的第一个世纪里，书籍的生产规模呈指数型增长，数量增长了近 30 倍。人类知识旋即进入快速积累期。

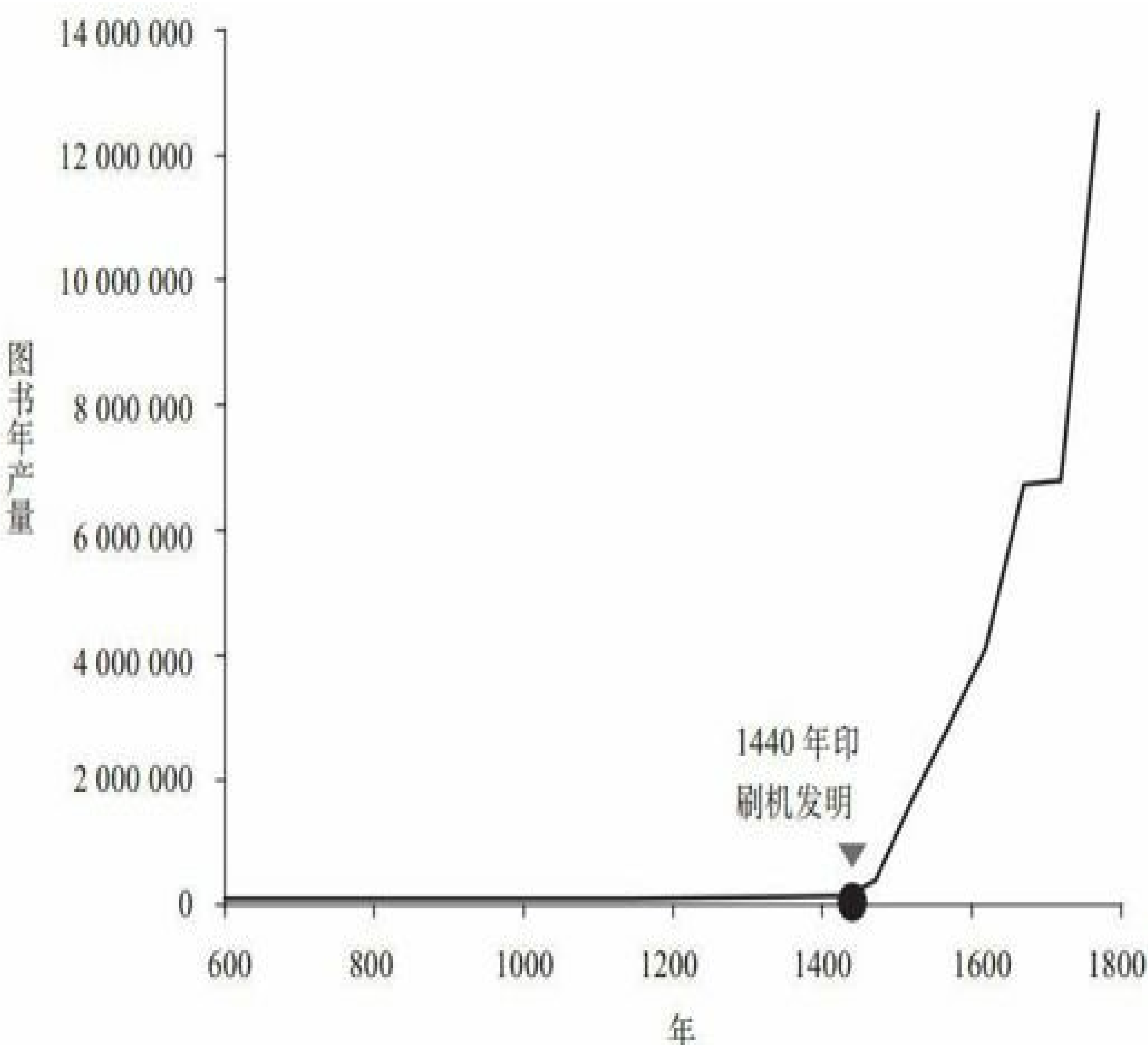


图 **0-1** 欧洲图书产量

然而，正如万维网建立初期那样，印刷机使用之初的信息质量也是参差不齐的。当时，人们只顾追逐眼前利益，印刷机几乎都用来印制那些质量较高的地图了，异端的宗教文章和一些伪科学文章也很快就占据了畅销书单的主要位置。印刷错误大量出现，那本被叫作《邪恶圣经》的书便是如此，这本书犯了史上最严重的印刷错误——《十诫》中的“不可奸淫”误印成“应当奸淫”。与此同时，人们一下子接触到大量的新思想，这难免会产生诸多混淆。信息的增长速度远远超过了人们处理信息和分辨信息的速度。共享信息的不断增长反而加速了民族和宗教的孤立进程，其速度之快不禁让人瞠目结舌。面对“过量的信息”时，我们会本能地进行筛选，选出喜欢的，忽略其他的，与同道中人为友，与意见相左之人为敌。

印刷机的早期使用者中，最狂热的要数那些传播福音的人了。马丁·路德的《九十五条论纲》还不是那么激进，而对其中类似情绪的争论已然不绝于耳了。正如《现代欧洲早期印刷革命》一书的作者伊丽莎白·爱森斯坦所写的那样，马丁·路德这些论纲的革命之处就在于，它们“没有一直钉在教堂的门上”。而是被古腾堡的印刷机复印了 30 多万次，即使按照今天的标准来看，这个印刷量也可算作巨大的成功了。

路德的新教改革所导致的教会分裂，很快使欧洲陷入了战争。1524~1648 年间，欧洲爆发了德国农民战争、施马尔卡尔登战争、八十年战争、三十年战争、法国宗教战争、爱尔兰南联邦战争、苏格兰内战和英国内战，其中许多战争几乎是同时爆发的。当然，1480 年出现的西班牙宗教法庭和 1508~1516 年出现的康布雷同盟也值得一提，尽管两者与新教思想的传播关系不大。单单一个三十年战争，德国人口就减少了 1/3，即使是与 20 世纪早期的第一次世界大战相比，17 世纪也称得上是史上最血腥的时代。

但就是在这样一个充满战乱的时代，印刷机却悄然推动着科学与文学的进步，所以人们才得以分享伽利略的科学思想，品味莎士比亚的舞台剧。

莎士比亚舞台剧的主题通常关乎人的命运，这一点与戏剧如出一辙。剧中人物的理想和命运之间的差距增添了这些舞台剧的悲剧色彩。莎士比亚时期，掌控自身命运看似已成为人们思想意识的一部分，但要做到这一点却很难，所以那些挑战命运的人总是遍体鳞伤，直至死去。

这一主题在莎士比亚的著名悲剧《朱利叶斯·恺撒》中得到了最生动的诠释。在这部戏的前半部分，恺撒接收到各种各样明显的警示，他称之为预言（比如“留心 3 月 15 日”），这些预言预示了他的加冕礼可能会演变为一场屠杀。恺撒当然不会在意那些预言，他始终自信地认为，

这些预言只是预示着别人的死亡，否则，他当时就会留意。结果，恺撒遇刺了。

莎士比亚借西塞罗的话警示我们，“(可是)人们照着自己的意思解释一切事物的原因，实际上却和这些事物本身的目的完全相反”。这句话对于所有正在对自己新发现的大量信息进行选择的人，都不失为一条好的建议。人们很难从干扰他们的噪声中分辨出有用的信号。数据展示给我们的通常都是我们想要的结果，而且我们通常也能确保这些数据令我们皆大欢喜。

然而，如果说《朱利叶斯·恺撒》这部戏剧中含有的宿命论、占卜术和迷信思想，是古代预言思想的开端，它同时也介绍了一种较为现代且较为激进的思想：我们可以对这些迹象进行解释，从中获益。在戏中，卡修斯说，“有时，人们可以掌控自己的命运”，他希望能劝服勃鲁托斯参与谋害恺撒的阴谋中。

于是，“人们可以掌控自己的命运”这一思想便广泛传播开来。“预言”和“预测”这两个词在今天大多数的情况下可以互换使用，然而在莎士比亚时代，它们却有着不同的含义。“预言”是指占卜者告诉你的话，而“预测”则更像是卡修斯的想法。

“预测”一词源于日耳曼语，而“预言”一词源自拉丁语。“预测”反映的是新教世俗思想，而不是神圣罗马帝国的理想世界。“预测”是指在不确定的条件下进行计划，这一行为需要谨慎、智慧和勤奋，更像我们今天所说的“预见”一词。

预言思想的神学含义是十分复杂的，但对于凡尘俗世中那些追逐利益的人来说，这些含义就不那么复杂了。预言思想的这些特质与那些新教徒的职业道德是密不可分的，马克斯·韦伯认为，资本主义的诞生和工业革命的开始与预言思想不无关联。“预测”与“进步”两个概念紧密相关。所有相关书籍中的所有信息都应有助于我们规划生活，都应成功地预见整个世界的发展历程。

几个世纪以来，引领“圣战”的新教徒们都在学习如何用自身积累的知识改变社会。工业革命主要始于新教国家，而且多半发生在那些言论自由的国家，因为在这些国家，宗教思想和科学思想可以自由传播，人们也不必对审查制度心存顾虑。

工业革命的重大意义难以尽数。纵观人类社会的历史进程，经济增长的速度曾经为年均 0.1%，这个增速足以匹配当时人口数量平缓增长的状况，但人均生活水平却没有得到任何显著提高。然后，经济形势突然出现了前所未有的进展，经济增速急剧超越了人口数量增长的速度，尽管偶尔也会出现全球金融危机，但这种高速增长的态势时至今日仍未

改变。

历史证明，印刷机引发的信息大爆炸为我们创造了一个好的世界，因为它仅用了 330 年的时间就为我们带来了不可尽数的好处，而与此同时有几百万人在欧洲战场上丧命。

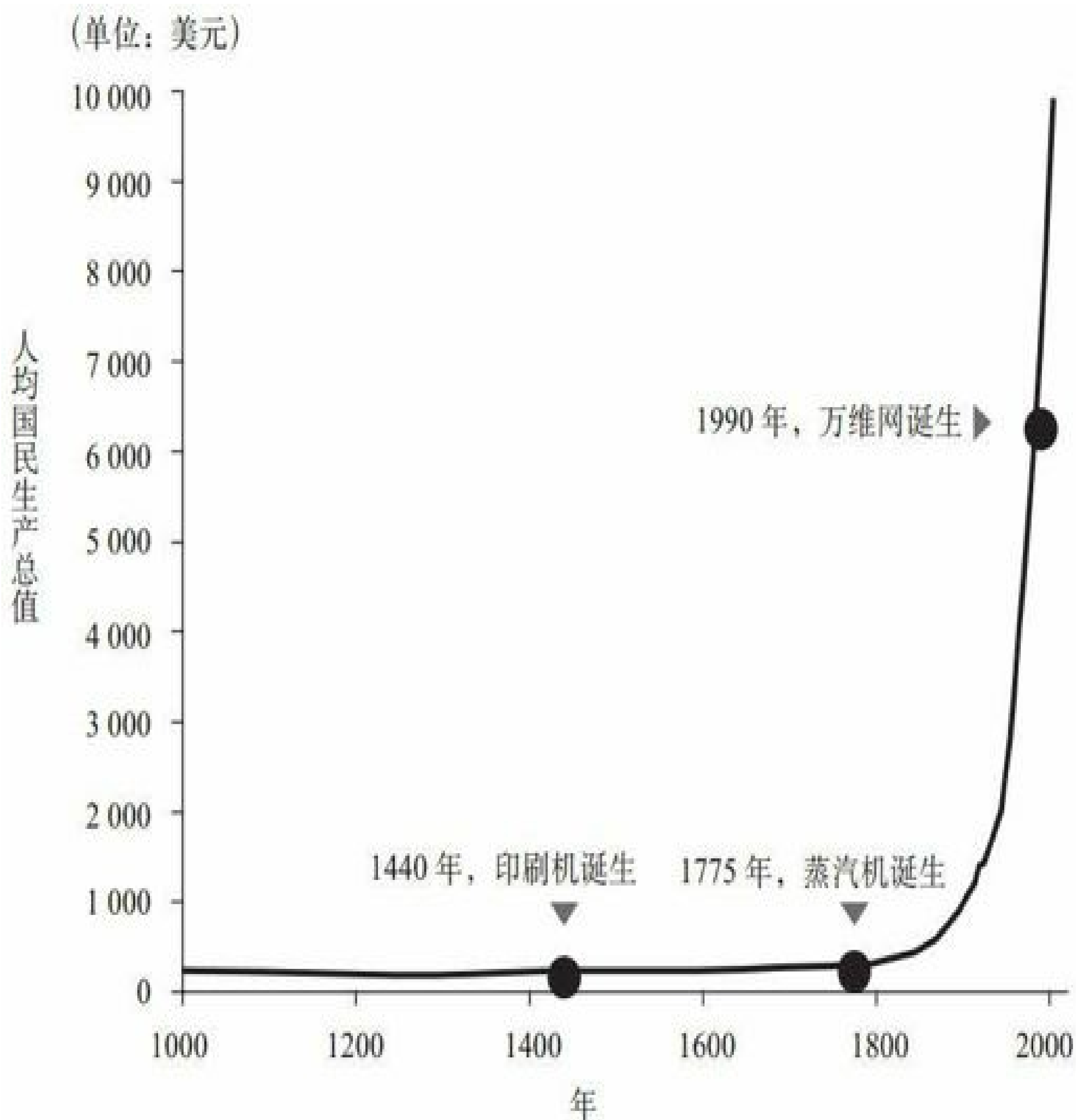


图 **0-2 1000-2010** 年间全球人均国民生产总值

生产力悖论

一旦信息增长的速度过快，而我们处理信息的能力尚且不足，情况

就很危险。过去 40 年的人类历史表明，把信息转变为有用的知识可能还需要很长时间，一不小心，我们就有可能倒退回去。

“信息时代”并不是特别新的术语，自 20 世纪 70 年代后期开始，这个术语便得到越来越广泛的使用。而与之相关的术语“计算机时代”使用得还要早一些，1970 年就有人用了。也就是在那个时候，计算机虽然没有在家庭中普及，但在许多实验室和学术机构中却越来越普及了。这一次，没有经过 300 年的时间，信息技术的增长就已经开始为人类社会创造实实在在的利益了，但是人们也等了 15~20 年才享受到这些利益。

正如美国经济学家、诺贝尔经济学奖得主保罗·克鲁格曼曾经对我讲的那样，20 世纪 70 年代是典型的“大量理论堆积如山，验证数据少得可怜”的时期。当时，人们已经开始用计算机制作地球模型，但是一段时间以后，人们发现这些模型太过粗劣，而且与实际不符，计算机可达到的精确度根本无法替代预测的准确度。这一时期，人们作过很多大胆的预测，涉及范围从经济学到流行病等各个领域，但是这些预测通常都不准确。比如，1971 年，人们声称可以预测出未来 10 年内的地震次数，而实际上，这在 40 年后的今天仍实现不了。

20 世纪七八十年代的计算机热非但未能推动经济和科学的发展，反而造成了两个领域生产力水平的短暂下降。经济学家将这种现象称为“生产力悖论”。罗伯特·索洛曾经在 1987 年写道：“计算机无处不在，可生产统计中却不见其身影。”1969~1982 年，美国经历了 4 次经济大衰退，直到 20 世纪 80 年代后期，美国经济才开始好转，而世界其他国家的经济状况则鲜有起色。

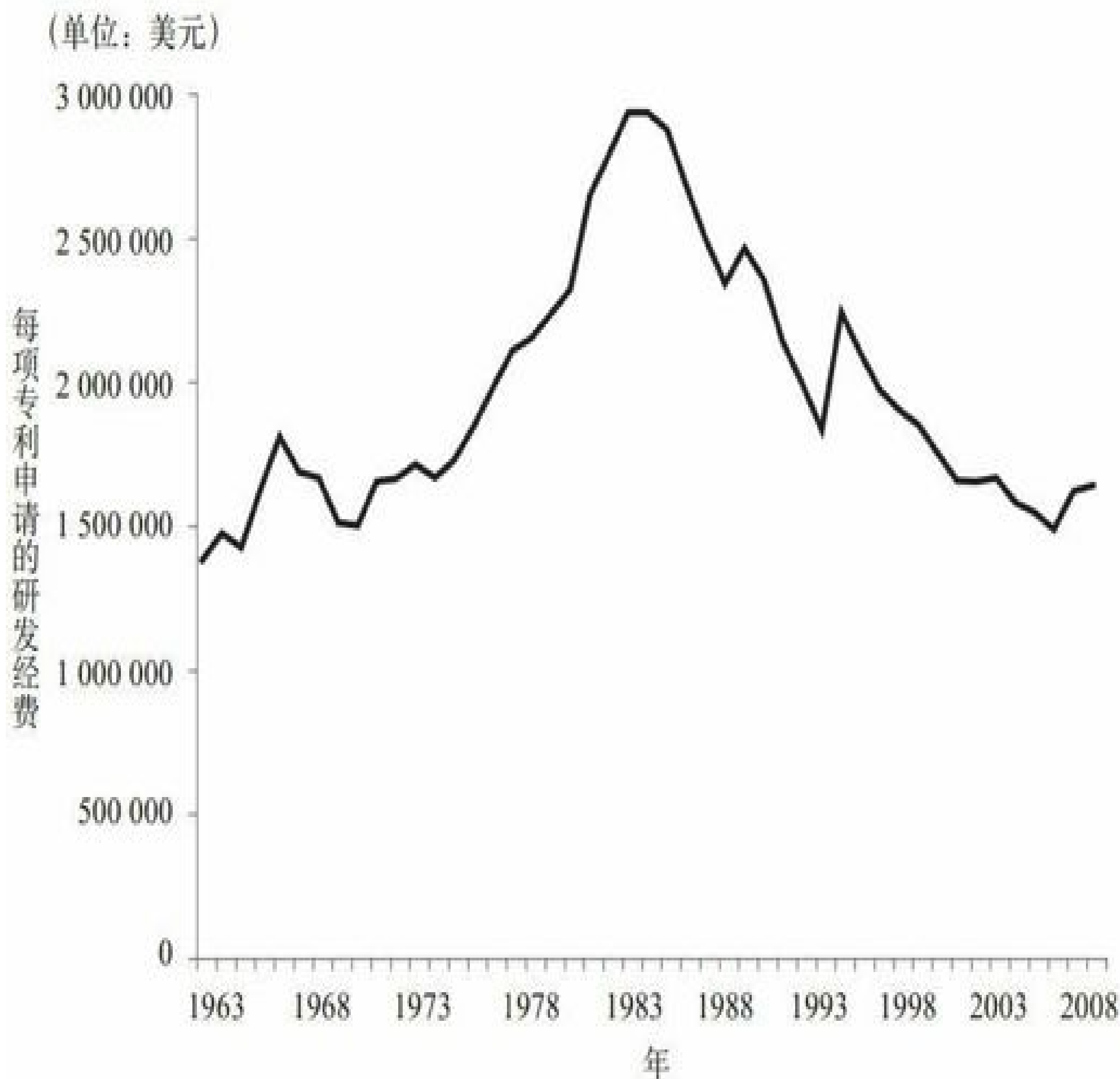


图 **0–3** 每项专利申请的研发经费

科学发展比经济发展更难判定，但科技进步有一大标志，即专利的数量，尤其是与研发投资相关的专利数量。如果一项发明的成本降低了，这就表明我们善于利用信息，并将其转变为知识。而如果发明的成本增加了，那就说明我们正在噪声中寻找信号，这无疑是在错误的方向上浪费时间。

20 世纪 60 年代，美国在每个发明者的每项专利上的耗资都接近 150 万美元（通货膨胀因素考虑在内）。这一数字在信息时代初期有增无减，1986 年更是成倍增长，最高达到 300 万美元。

因为越来越注重新技术带给我们的实惠，20 世纪 90 年代我们又一

次提升了科学研究的生产力。各项研究不再硬钻“牛角尖”，计算机开始被用于改善人们的日常生活，促进经济发展。通常情况下，许多预言从长远角度看算作进步，而从短期角度看则成了倒退；而许多从长远来看似乎可以预知的事情，同时也会妨碍我们进行完美的计划。

“大数据”的承诺与陷阱

时下最流行的术语要数“大数据”了。根据国际商业机器公司（IBM）估计的数据来看，现在我们每天生成的数据高达 250 兆亿个字节，超过过去两年里生成的数据总量的 90%。

信息的指数型增长有时被人们视为万灵药，就好比 20 世纪 70 年代出现的计算机一样。《连线》杂志的前主编克里斯·安德森曾经在 2008 年的一篇文章中说：“数量庞大的数据会使人们不再需要理论，甚至不再需要科学的方法。”

本书着重介绍了前沿科学和高端技术，我认为其内容是积极乐观的，却被严重曲解了。虽然那些数字不能为自己辩护，但我们却可以作为数字的发言人，赋予它们意义。这就好比对恺撒密码解码一样，我们可能会以对自己有利的方式对这些数据进行分析和解释，而这些方式很可能与这些数据（所代表）的客观现实不相吻合。数据驱动预测机制可能会成功，也可能会失败。一旦我们否认数据处理过程中存在着主观因素，失败的概率就会增加。要提高数据分析的质量，首先要对我们自身提出更高的要求。如果对我的情况不甚了解，你可能就会对前面的提法颇感意外。我在数据和统计学领域还算小有建树，曾经据此做出不少成功的预测。2003 年，由于厌倦了咨询工作，我设计了一个名为“PECOTA”的系统，主要用来预测美国职业棒球联盟球员的各项数据。这个系统有很多创新点——其预测是概率性的，比方说其中为每位球员都列出了一系列可能出现的结果——当我们将这些预测结果与比赛系统给出的结果进行比较时，发现这套系统的性能更加优越。2008 年，我建立了“FiveThirtyEight”（538 网站，因 538 张选举人总票数得名）网站，试图对即将举行的美国总统大选进行预测。该网站对两位美国总统候选人在美国 50 个州中的竞选结果进行了预测，结果命中 49 次竞选，只有 1 次失手。另外，我的网站还预测出美国参议院选举的 35 个席位归属。

美国总统大选过后，很多出版商找到我，希望我能为《点球成金》和《魔鬼经济学》这类关于小人物征服大世界的书籍估个价。本书涵盖很多行业和领域，从金融领域，到国家安全，对这些行业和领域中的数据驱动预测机制进行了调查。

在 4 年时间里，我曾与十几个领域中的 100 多位专家交流过，读过数百篇期刊文章和论文，为了实地调查，我跑遍了从拉斯韦加斯到哥本哈根的许多地方，却发现“大数据”时代的预测活动发展得并不顺利。我的成功也只是因为我在某种程度上比较幸运，一是尽管出现了本书中提到的一些错误，但还是取得了成功；二是选对了调查案例。

本书提到了几个值得研究的例子，集人类判断与计算机功能为一体的天气预报就是其中之一。虽然气象学家的名声不好，可是他们也取得过显著的进步，比如他们预测飓风着陆位置的准确度比 25 年前提高了 3 倍。与此同时，我还拜访过一些在赌城拉斯韦加斯轰动一时的扑克牌玩家和（体育赛事）赌徒。

但是，这些预测成功的案例必然是建立在一系列失败案例的基础之上的。

如果让我们用一个特点来定义美国人——一个令其与众不同的特质——那就是美国人对卡修斯精神的信仰：我们的命运由我们自己主宰。一些宗教叛逆者迎着工业革命的曙光建立了美国，他们认为自由流动的思想不仅有助于传播其宗教信仰，也有助于传播科学和商业贸易，“作为一个民族，我们的智慧、我们的勤劳、我们的傲慢和急躁、我们所有的强项和弱项，都源自我们那不可动摇的信念，那就是我们要为自己做主”。

新千年给美国人带来的是噩梦般的开始。我们没有预测到“9·11”恐怖袭击事件，而这一惨剧的出现并非因为我们的信息匮乏。正如 60 年前的“珍珠港事件”一样，其实所有的信号都在那里，只是我们没能将它们联系起来。因为对恐怖分子可能会有的举动不够了解，所以我们对那些数据视而不见，不知道大难将至。

近期，对全球金融危机的预测也总是失败。我们天真地相信各种（预测）模式，却没有认识到这些模式在我们进行假设选择时根本不堪一击，因此总会带来惨痛的后果。在日常生活中，我发现尽管人们也在努力尝试，却仍然无法提早预测出经济衰退。幸好在控制通胀方面，我们已经取得长足进步，否则那些经济决策者就只能“盲目飞行”了。

与 20 世纪 70 年代一样，近来人们十分热衷于对地震进行预测，其中大部分高度依赖数学方法和数据处理技术。但是，这些预测只是假想一些从未发生过的地震，对真正发生的那些地震却没有预测到。福岛核反应堆的设计可以抵抗 8.6 级地震，因为一些地震学家称不可能发生更高级别的地震。但是，2011 年 3 月日本却发生了 9.1 级的特大地震。

错误地预测整个学科的发展常会危及整个社会。以生物医药学的研究为例。2005 年，一位土生土长的雅典人，医学研究者约翰·P·埃尼迪

斯，发表了一篇颇具争议性的论文，题为“为什么大多数发表的研究成果都是骗人的”。该文对那些行业期刊中刊载的积极的研究成果进行了研究（这些成果认为那些在实验室实验中得到验证的医学假设堪称成功预测），认为大多数成果在实际生活中很可能是毫无用处的。德国拜耳制药公司最近证实了埃尼迪斯的这个推断，他们通过实验亲自对那些医学期刊中提到的积极研究成果进行验证，但发现其中近 2/3 的医学假设根本不能成立。

这些大数据终将推动社会进步，至于这种进步的速度有多快，或者进步的同时是否还会倒退，这些都取决于我们自己。

为何未来使我们震惊？

人类并没有多少天生的防御能力，人类的速度没有多么快，身体也没有多么强壮；人类没有利爪和尖牙，也没有护身的硬壳；人类不能喷出毒液，不能伪装自己，也不能飞翔。我们之所以能生存下来，是因为我们运用了智慧。我们的思维很敏捷，我们能够敏锐地感知事物的模式，对机遇与威胁迅速地做出反应。

“人类比其他动物更需要发现模式”，麻省理工学院的神经系统学家托马索·波吉奥对我说，他的研究领域是人脑对信息的处理模式。“在复杂情境中识别物体的能力是一种概括能力。一个新生儿就能识别人脸的基本模式，这种能力是进化而来的，是人类生来就有的，并非后天习得的。”

但波吉奥认为，问题在于这些进化来的本能有时会让我们去寻找原本不存在的模式，“人们一直都在努力从随机噪声（即无规律的状况）中发现模式”。

人脑能力非凡，其信息存储量或高达 3 千兆字节。然而，据 IBM 公司称，人脑的存储量不过是全球每天所产生信息量的百万分之一而已。因此，我们对自己记忆的信息一定要精心挑选才行。

在 1970 年出版的《未来的冲击》一书中，未来学大师阿尔文·托夫勒对他所说的“信息超负荷”的一些后果进行了预测。他认为，尽管世界本身正走向分化，变得更加复杂，但人类仍会以坚持自身看法的方式使这个世界变得简单，这便是我们的防御机制。

我们的生物本能有时会对这个信息丰富的世界难以适从。所以，我们需要积极努力，坚持自己所持有的看法，这样才有可能将重返信息负荷状态的可能性降到最低，甚至消除这种可能性。

印刷机诞生后，信息超负荷催生了更深层的宗教主义。现在，那些不同的宗教思想可以通过更多的信息、信念和“证据”得到证明，而且更

难以容忍反对意见。同样的情况似乎到今天都一直存在。在托夫勒刚刚开始写《未来的冲击》这本书时，美国的党派政见分歧开始加剧，这种分歧也许会随着互联网的出现愈演愈烈。

不同的党派政见颠覆了“信息越多，就越靠近真相”这一信条。《自然》杂志上最近刊登的一项研究发现，几大政党对全球变暖的问题了解得越多，他们之间达成共识的可能性就越小。

同时，如果信息的数量以每天 250 兆亿字节的速度增长，其中有用的信息肯定接近于零。大部分信息都只是噪声而已，而且噪声的增长速度要比信号快得多。有太多假设需要验证，有太多数据需要发掘，但客观事实的数量却是个相对恒量。

印刷机改变了我们犯错误的方式，从前经常出现的抄写错误越来越少。然而一旦出现错误，这个错误就会被多次复制，《邪恶圣经》里出现的错误就是这种情况。

像万维网这类复杂的系统有这样一个特点，它们不像那些相对简单的系统那么容易出错，但一旦出错，必定是要命的大错。在信息宣传方面，资本主义和互联网都非常高效，这就使得好坏两种消息的广泛传播成为可能，而且坏消息也许会造成更大的影响。金融危机爆发之前，这一体系正处于高度负债经营的状态，那些信誉评级机构的预测模式一旦做出错误的假设，哪怕只是一个错误的假设，都极有可能摧毁全球的金融系统。

制定规章制度是解决这些问题的途径之一，但是我怀疑这种制度会成为我们逃避自身问题的借口。我们要停止对事物进行预测的做法，并且承认我们的预言有问题。我们喜欢对事物做出预测，而我们的预言却总是出错。

预测与贝叶斯定理

如果说预测是本书的中心问题，那么它同时也是这一问题的解决方案。

预测在我们的生活中必不可少。每次我们选择工作方法、决定是否要与对方展开恋情或是未雨绸缪地将钱存起来，都是在对未来将要发生的事情进行预测，也是在对我们计划的成功概率进行预测。

不是所有的日常问题都需要费心思量，我们作每个决定的时间是极为有限的。尽管如此，我们每天还是要进行许多预测，只是有时没有意识到而已。

正是出于这个原因，本书将预测视为人们共同的事业，而不仅仅是一些顶级专家或相关从业者的职责。取笑那些专家的预测失误总是乐趣

无穷，而在幸灾乐祸的同时，我们自己也要多加小心。如果说我们的预测不会比那些专家更糟，只不过是欲贬先颂罢了。

然而不可否认，预测在科学领域中的确扮演着重要角色。你们中有些人也许对我此前提到的一个前提感觉不舒服，现在我来澄清一下这个前提：我们永远都不可能做出完全客观的预测，因为这些预测总会带有主观色彩。

但是，本书对“根本不存在客观真理”这一虚无缥缈的说法完全不赞同。相反的，本书认为要做出准确的预测，首要的前提就是坚信客观真理的存在，并且执着地追寻它。而预测者的另一个承诺，就是要认识到他无法穷尽对客观真理的认知。

预测之所以重要，是因为它连接着主观世界与客观现实。科学哲学大师卡尔·波普尔早就意识到了这一点。对他来说，假设并不科学，可证伪的假设才是科学的。这就意味着在真实世界里，假设可以通过预测得到检验。

令我们裹足不前的是，经过验证的那些为数不多的想法的实际作用并不大，而且许多想法未经过检验，或者根本就无法检验。在经济领域中，验证失业率预测的准确性要比验证刺激消费政策的效果的论断容易得多。

我不会像波普尔那样，断言这样的理论不够科学或者没有价值。事实是，少数几个可以检验的理论得出的结果很糟糕，这就表明我们没有检验的那些想法中有很大一部分是大错特错的。毫无疑问，我们生活在幻想中却不自知。

但解决方法还是有的，这种方法不依靠不成熟的政策—尤其是当我意识到出现这个问题在很大程度上是因为美国的政体有问题时—它需要的是态度上的转变。

这一态度会通过贝叶斯定理得到体现，我在第八章中会讲到这一定理。贝叶斯定理名义上是一个数学公式，但其内涵却远远超出公式的范畴。这一定理表明，我们必须从不同角度去思考我们的想法，以不同的方式检验我们的想法。我们要坦然面对各种可能性和不确定性，更加周详地考虑我们对一个问题的假设和看法。

本书可以粗略地分为两部分。前 7 章的内容分析预测的问题，后 6 章的内容对贝叶斯定理进行了探讨和应用。

每一章都围绕着一个特定的主题展开，并且对其进行深入分析。不可否认，这是一本复杂详尽的书，一方面因为细节通常可以揭示关键问题；另一方面，在我看来，专注于某一问题的研究总会得到比概括研究更深刻、更独到的见解。

我选择的话题通常是公众共享的信息，预测的案例很少依据私人信息选择（公司利用客户记录预测新产品的需求量）。我更倾向于那些大家可以一起挖掘真相的话题，你不需要只听我的一面之词。

本书的阅读路径

本书涵盖了自然科学、社会科学以及体育运动等各个领域的诸多实例。书中将许多相对通俗易懂的案例改造成需要稍微加以巧妙处理的案例。这些案例对成功的预测与失败的预测进行了清楚的划分。

第一章至第三章讨论了近期出现的金融危机大背景下的失败的预测，预测中有的很准确，有的则很离谱。预测会让你想到关于预测的最基本的问题：在应用数据时，我们怎样才能做出客观的判断呢？关于市场竞争，什么时候可以做出更准确的预测呢？当我们意识到未来与过去不一样时，又如何在两者之间进行协调呢？

第四章至第七章重点讲述动态系统：地球大气层的动态会带来天气的变化，构造板块的运动会引发地震，复杂的人类活动会影响美国经济，还有传染病的传播、扩散。最优秀的科学家正在研究这一系列的因果关系，然而各类动态系统使得预测变得更加困难，这些领域的预测活动总是不尽如人意。

第八章至第十章主要探讨解决方案。首先介绍了一个体育赌徒，他在运用贝叶斯定理时比许多经济学家或科学家还专业，之后讲到扑克牌。体育和游戏都遵守规则，这一点是检验预测技巧的实验所必需的。这几章帮助我们更好地理解什么是不可测性和不确定性，让我们深刻理解如何把信息转变为知识。

然而，贝叶斯定理也可以被用来解决更多现存的问题。第十一章至第十三章讲到了 3 个问题：金融市场泡沫、全球变暖和恐怖主义。这 3 个问题对那些预言家和美国社会而言，都很棘手，但如果接受挑战，我们就能使我们的国家、我们的经济和我们的星球更加安全。

自从有了印刷机，我们的世界已经经历了太多。信息不再那么稀有，我们拥有的信息太多，甚至多到无从下手，但有用的信息却寥寥无几。我们主观地、有选择地看待信息，但对信息的曲解却关注不够。我们以为自己需要信息，但其实我们真正需要的是知识。

信号是真相，噪声却使我们离真相越来越远。本书为你呈现的就是信号和噪声的故事。

第一章
预测失败的灾难性后果

2008 年 10 月 23 日，股票价格较前 5 周暴跌了约 30%，美国股市崩盘。许多如雷曼兄弟公司一样曾备受推崇的公司纷纷宣布破产。在拉斯韦加斯，多家公司市值缩水 40%。失业率飙升。金融机构破产，数千亿美元流失。民意测验显示，当时美国民众对美国政府的信心已降至最低点。此时距美国的总统大选还有不到两周的时间。

美国总统大选前通常很平静的国会，这次却为救市法案站在了风口浪尖上。这项刚刚通过的法案注定不受欢迎，所以它需要尽力地给公众留下维护正义的印象。美国众议院监督委员会已经约见了标准普尔、穆迪以及惠誉国际这三大信用评级机构的高管。这些评级机构受到指控，因为它们评估数万亿住房抵押贷款支持证券存在拖欠的可能性。说得婉转些，这些评级机构的判断出现了失误。

一场错误预测引发的悲剧

2007 年爆发的危机常被视为政治体制和金融机构的失败。这次危机显然是一次大规模的经济衰退。直到 2011 年，也就是这次经济大萧条正式开始的 4 年之后，美国的经济增长仍然低于其生产潜力 8 000 亿美元之多。

然而，我十分确信，我们更应该将这次金融危机视为决策失败的结果和预测严重失准的结果。这些失败的预测影响面之广，贯穿了这次金融危机的全过程；涉及人数之多——从抵押贷款经纪人到美国白宫的官员，几乎所有人的预测都是错误的。

最失败的预测通常有很多共同点，即我们只关注那些符合我们对这个世界的期许的信息，而不在乎其真实性。对于那些最难测定的风险，即使它们严重威胁到我们的幸福生活，我们也会对其视而不见。我们所作的各种预估和假设要比现实状况简单得多。即使在解决问题时绕不开不确定性，我们也会刻意地回避它。要想了解此次金融危机的关键点，我们首先得找出所有失败的预测中错得最离谱的那个，因为所有的错误都因这个预测而起。

正常来讲，评级机构会将 3A 评级给予全球范围内少数偿还能力高的政府和运营状况极佳的公司。而实际上，数以千计的住房抵押贷款支持证券和金融工具居然也都拥有 3A 评级，于是便滋生了投资者靠别人的住房贷款违约生财的情况，他们赌的是他人违约的概率。这些评级机

构给出的评级结果在很大程度上会被人们视为预测：对一项贷款拖欠的可能性作出预估。例如，标准普尔评级机构公布某个担保债务凭证为 3A 级，这便意味着这项债务在未来 5 年被拖欠的可能性只有 0.12%——概率只有 1/850——与其他 3A 级企业的债券一样安全，甚至比美国国债还要安全。这些评级机构从来不进行曲线评级。

但事实上，根据标准普尔评级机构的内部数据来看，3A 级担保债务凭证的违约率约为 28%（一些独立统计机构得到更高的比例），这就意味着这些担保债务凭证的实际违约率比标准普尔评级机构先前预估的高出 200 多倍。这简直可以算作最失败的预测了：原本被认定为绝对可信的数千亿美元的投资，事实上却是毫无安全性可言。这就好比天气预报预测某天是一个温度高达 30 摄氏度（华氏 86 度）的大晴天，结果那天却袭来了暴风雪。

如果仅仅进行了一次错误的预测，还可以就此进行解释。其中一条途径就是将这次错误归因于外部环境，也就是我们所说的“运气不佳”，有时这种解释显得名正言顺，甚至让人觉得它就是实际情况。美国国家气象局预报明天天高气爽的概率为 90%，结果却下雨了，害得你打高尔夫的计划泡汤了，而你却不可能对国家气象局动怒。几十年来的历史数据显示，当天气服务信息预测降雨概率为 10%，当天下雨的可能性的确占到很长一段时期以来的 10%。

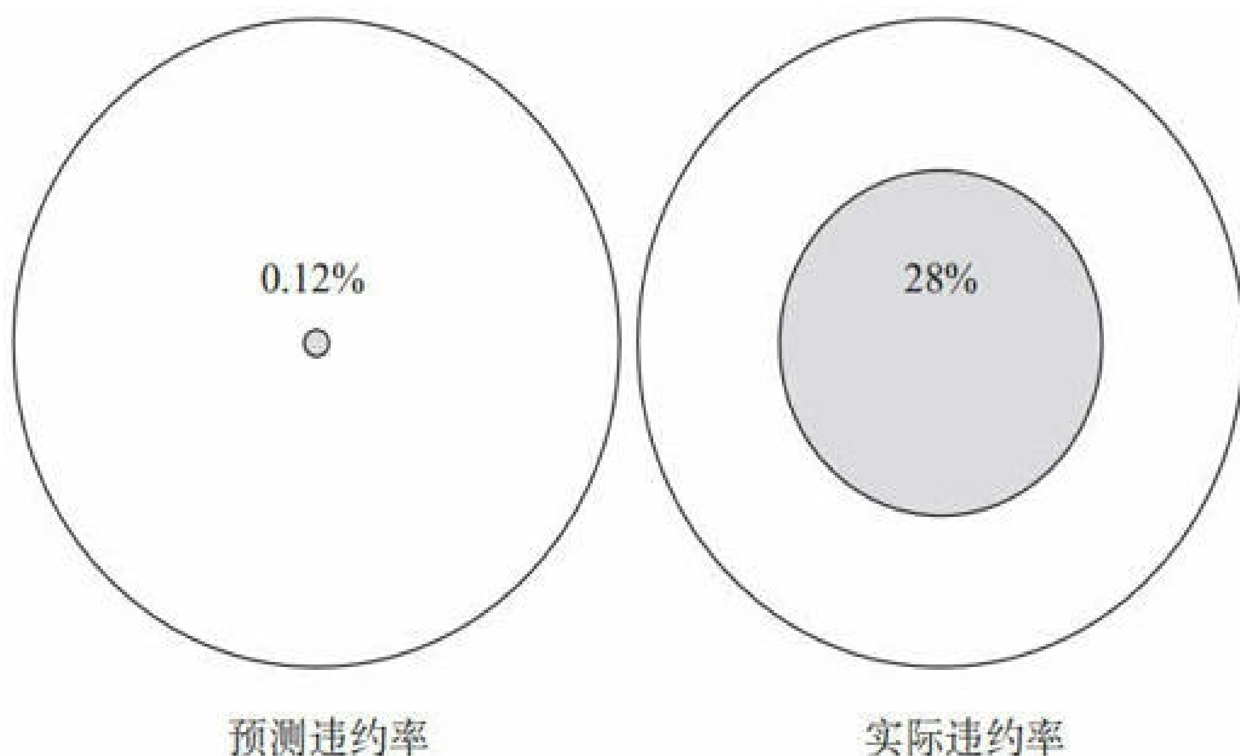


图 **1–13A** 级担保债务凭证的预测违约率与实际违约率的 **5** 年数据比较

然而，如果预测者从未有过成功的预测，或犯下严重的错误，“运气不佳”这个理由就站不住脚了。因为失败的预测和客观世界无关，而往往取决于预测者的主观方法。

先前提到的担保债务凭证的例子，那些评级机构的预测根本就无据可依：这些证券是全新的证券，关于它们的记录几乎为零；标准普尔评级机构给出的违约率并非由历史数据推导而来，而是在错误的统计模型基础上做出的假设。这一错误评级带来了严重的后果——所谓 3A 级，违约率竟比理论中高出了 200 多倍。

评级机构应该做出的补救是，承认那些模型是有缺陷的，承认自己犯了错误。但在美国国会听证会上，评级机构的负责人都以“运气不佳”为托词意图逃避责任，谴责房地产泡沫带来了此次意外。

标准普尔公司总裁德文·夏尔马于 2008 年 10 月曾对美国国会说过：“对房地产市场和抵押信贷市场的急速衰退感到吃惊的不只是标准普尔公司，实际上，无论房主、投资者、金融机构、评级公司还是监管人，谁都没有预料到会出现这种情况。”

谁都没有想到会出现这种情况。无法证明自己的清白，那就承认自己的无知：一旦预测失败，人们通常会将这种手段作为自己的第一道防线。但是，夏尔马的上述说法其实是在撒谎，这种说法就和“我和那个女人根本没有发生性关系”或“我从没有服用过类固醇”之类的说法一样，不过是美国国会听证会上的惯用伎俩。

然而，这次房地产泡沫真正令人瞩目的地方就在于，很多人此前已经想到会出现这种情况，并且很笃定地将这种情况提前告知给人们。早在 2000 年的时候，耶鲁大学经济学教授罗伯特·希勒在其著作《非理性繁荣》中就注意到了房地产泡沫的苗头。英国经济政策研究中心的经济学家迪安·贝克于 2002 年 8 月也提到房地产泡沫问题。一向以文风保守著称的《经济学人》杂志，在 2005 年 6 月刊发了一位通讯记者的相关文章，这位记者在文中也提到了这次“史上最大的泡沫”。诺贝尔经济学奖获得者保罗·克鲁格曼在 2005 年 8 月也撰文记述了这次房地产泡沫及其必然后果。后来，克鲁格曼告诉我：“房地产泡沫是内置在经济系统中的。房市崩盘并非黑天鹅，而是房间里的大象，看上去显而易见却总是被人们忽略。”

其实，普通的美国民众对这个问题也很关切。2004 年 1 月至 2005 年夏天，在短短一年半的时间里，“房地产泡沫”这一词条的谷歌搜索量迅速增加了近 10 倍。在房价飙升速度最快的那些州，比如加利福尼亚州，人们对该词条的关注度最高。而这些州的房价也将经历最大幅度的降价风暴。事实上，关于“房地产泡沫”这一话题的讨论当时迅速展开，

引起了人们的广泛关注。2001 年的新闻报道中出现“房地产泡沫”这一词条的只有 8 条新闻，而到了 2005 年，相关新闻的数量已经跃升至 3 447 条，这一词条在知名的报纸期刊上出现的频率也高达日均 10 次。

然而，这些负责预测金融市场风险的评级机构竟然说它们没有注意到金融市场存在风险。这下你该明白，这些机构似乎已经把这种说辞当作它们的最佳防线了。它们对自己所作预测中存在的问题，可谓讳莫如深。

他们只是不想让“音乐”停下来罢了

我曾就本章内容与多位经济学家和投资者交流过，他们对那些评级机构的工作全都不满意，但是他们认为导致这些机构预测错误的原因可以分为两类，即因贪婪而犯错和因无知而犯错。可问题是，他们真的那么无知吗？

对于这个问题，朱尔斯·克罗尔也许有资格做出评判，因为他本人就经营着一家评级机构——“克罗尔债券评级公司”。这家公司成立于 2009 年，我是 2011 年在克罗尔位于纽约的办公室里见到他的，当时他的公司刚刚发布了一个项目的初评结果——对弗吉尼亚州阿灵顿一家大型购物中心的建筑商的抵押贷款进行评级。

克罗尔指责大多数评级机构缺乏“监管”。“监管”这个词从克罗尔嘴里说出来颇具讽刺意味，因为在涉足评级行业之前，他就已经因为创建“克罗尔风险顾问公司”而小有名气了（虽然为人低调，赚钱却很高调，且赚得盆满钵满）。这家风险顾问公司相当于一家侦探公司，专门侦查各类公司的欺骗行径。这家公司知道如何识破骗局——比方说，一群绑架者劫持了身价过亿的对冲基金巨头，但这群绑匪用被绑架者的信用卡买了比萨，这就相当于把自己暴露了。我见到克罗尔时，他已经 69 岁高龄，但他的侦探直觉仍十分敏锐，在他着手调查那些评级机构的运营情况时，这些直觉便被充分地调动起来了。

克罗尔告诉我，“监管”是评级行业的专业术语，指评级公司有义务将所了解的信息如实告知广大投资人。评级机构每月都会收到一份新的数据记录，里面记录着大量数据，比方说抵押贷款的拖欠与支付情况等。这些数据可以提供早期预警——情况正朝着好的方向发展，还是越来越糟？全世界都等着那些评级公司发布这些信息。

换句话说，那些评级公司本该最先察觉到房地产市场的问题，因为它们比别人掌握了更多的有效信息——数以千计的借款人是否及时还款的第一手资料。但这些评级机构一直都没有大批量地下调住房抵押贷款证券的信用级别，直到 2007 年这些问题凸显出来，房屋止赎率也已经

上涨了 1 倍，此时评级机构才不得不着手处理这些问题。

克罗尔对我说："评级机构的人并不傻，他们对后果一清二楚。我想他们只是不想让音乐停止罢了。"

"克罗尔证券评级公司"是十大"美国认可的统计评级机构"之一，由美国证券交易管理委员会许可，为负债性有价证券进行评级。但前 10 名中的另外 3 家评级机构穆迪、标准普尔和惠誉几乎瓜分了全部市场份额；其中标普和穆迪曾分别为金融危机前发行的 97%的担保债务凭证进行过评级。

标准普尔和穆迪评级公司能占据如此大的市场份额，其中一个原因就是两大巨头长期以来一直是利益集团的一分子。两者皆为合法寡头垄断链条的一环，政府在这一行业中实施准入限制。与此同时，标准普尔和穆迪的积极评价往往是在大额养老金的各项相关法规的允准下做出的——在购买债券之前，其中 2/3 的养老金法规规定债券销售前，必须有标准普尔或穆迪两者同时对债券进行评级。

尽管标准普尔和穆迪的员工收入与华尔街那些公司的员工不可同日而语，但两大公司利用其优势地位获得的特别收益还是相当可观的。1997~2007 年这 10 年间，穆迪公司来自所谓结构性融资评估的收入就增加了 800 多倍。此类评估也是公司在泡沫经济时期的主要业务，这些业务使得穆迪在房地产泡沫期内连续 5 年占据着标准普尔 500 家上市公司的利润榜首位。（2010 年，即便当时房地产泡沫已经破碎，各大评级机构的问题已经凸显，穆迪仍旧保持着 25%的盈利率。）

只要新的担保债务凭证源源不断地发行，大笔利润就唾手可得。投资者事发前无从知晓评级的准确性，评级公司因此没有比拼服务质量的热情。穆迪的总裁麦克丹尼尔就曾明确地告知董事会，评级质量是对公司利润影响最小的因素。

评级机构的利润方程很简单。每完成一次评级任务，评级公司就会从证券发行人那里得到一份报酬，债权抵押证券发行得越多，来自证券发行人的利润也就越大。通过融合不同类型的抵押证券，或由原类型衍生出新类型，债权抵押证券其实可以无限量发行，而这些评级机构又不会错过任何评估的机会。后来，政府的一次调查公开了穆迪两名老员工的即时信息交流记录，其中一位员工说，即使一种证券是"一群牛设计的"，穆迪也愿意评估它。在有些情况下，一些评级机构甚至还会协助证券发行者操控评估结果。在所谓的透明机制下，标准普尔评级机构为证券发行者提供评级软件的副本，这既方便了他们了解不良贷款的具体数额，确定再投放量，同时还能维持评级结果。

房地产泡沫的出现及其破灭，将会断送评级机构唾手可得的财路。

对威胁自己的风险，人类拥有一种超凡的能力，那就是对其视而不见，好像这么做，风险就会自动消失。如此看来，德文·夏尔马在听证会上的陈述似乎是合乎情理的，即便其他评级公司注意到房地产泡沫现象，但标准普尔公司却很可能真的忽略了这一信息。

然而事实上，标准普尔公司十分清楚可能会出现房地产泡沫问题，但它们给出的结论却是，这次泡沫没什么大不了的。标准普尔评级机构的发言人凯瑟琳·马茜斯曾向我提供过一份备忘录，其中详细地记录了该公司在 2005 年做的一次模拟测评，这次测评预测出，两年内美国的房价将下降 20%，这个降幅与 2006~2008 年间房价的实际降幅——30%——已经相当接近了。这份备忘录认为，标准普尔现有的各种预测模式可以有效地“捕捉到发生经济低迷风险的信号”，由这些模式评定出的高级别证券可以“经受住房价下跌，且不会发生信用降级”。

从某些方面来讲，上述情况比评级公司完全忽略“房地产泡沫”这一因素更令人担忧。我将会在本书后面的章节中探讨那些“未知的秘密”——那些我们从未意识到的风险的危险性。我们以为自己可以控制很多风险，但结果并非如此，也许这才是更大的威胁。在这些情况下，评级机构的盲目自信不仅是在愚弄自己，还会殃及他人。标准普尔公司的案例表明，这种自以为是的做法会伤害整个金融体系。道格拉斯·亚当斯在其著作《银河系漫游指南》一书中写道：“可能会出错的事情和完全不可能出错的事情之间的主要差别在于，一旦这件完全不可能出错的事情最终出了错，这个错误往往无法挽回或者根本无法补救。”

既然拥有科学、精确的预警信息，评级机构的预测模式为何还会错误百出，预测水准如此低下呢？

评级机构为什么会犯下致命的错误？

只有深入挖掘，我们才能发现问题的根源。要找到这个答案，我们需要对类似担保债务凭证这类金融工具的构成作进一步了解，还要对不确定性与风险之间的差别有所了解。

担保债务凭证是抵押债务的集合，会被分配到不同的资产池，或划分为不同的“等级”。有些资产池风险很大，有些资产池则很安全。我的朋友阿尼尔·卡什在芝加哥大学教授金融危机这门课，他想出一个简化了的担保债务凭证的案例，我在这里就用他的案例来解释这个问题。

假设你有一组抵押贷款，由 5 项贷款组成，假设其中每一项都存在5%的违约率。依据抵押贷款的不同级别，你可以设置风险渐次增加的多个赌注。其中最安全的赌注，我们姑且称为α池，在这里当 5 项抵押贷款全都违约时才会输掉赌局。风险最大的赌注为ε池，这里任何一项

贷款违约都会被套牢。其他资产池的风险以此类推。

然而，投资者更愿意将赌注押在ε池，而不是α池，这是为什么呢？原因很简单，ε池可以以较低的价格抑制较高的风险。如果你是一个不愿意承担风险的投资人，比如用养老金来购买证券，你的投资原则就不允许你投资评级很低的证券。如果要购买证券，你一定会从评级为 3A 级的α池中购买。

α池中的证券由 5 项违约率只有 5%的抵押贷款组成。只有 5 项贷款全都违约你才会赔本。这种情况下还会有风险吗？

实际上，这个问题并不简单，而问题也正在于此。不同的假设或估算会产生完全不同的结果。如果假设错误，整个数据模式都会谬以千里。

一种假设将每一项抵押贷款都视为独立的，在这种情况下风险是多元化的：如果克利夫兰的某位木匠拖欠贷款，这件事与丹佛的某位牙医是否拖欠贷款没有任何关系，那么在这种情况下，你赔本的风险就格外低，如同连续 5 次掷骰子都掷出两点的概率一样低。确切地说，这种情况出现的可能性是 5%的五次方，概率低到 320 万分之一。评级机构会宣称一组平均信用等级为 B+的次级抵押贷款属于这种假设的多元化奇迹——这种等级的贷款通常情况下意味着违约率会超过 20%，但放在一个资产池中，其违约率则趋近于零。

与上述假设截然相反的另一种假设是，5 项贷款之间并非完全独立，而是彼此休戚相关，要么 5 项同时违约，要么都不违约。这时就用不着再分别掷 5 次骰子，你其实是把赌注押在了一局上。你有 5%的概率押中两点，一旦押中，则 5 项贷款全部违约，这个概率比你最初设想的风险高出 160 000 倍。

表 **1–1** 简化的担保债务凭证结构

赌局失败的概率				
赌池	规则	各项贷款的违约率紧密相关的概率	各项贷款的违约率不相关联的概率	风险倍数
α池	5 项抵押贷款都违约为输，其他情况为赢	0.000 03%	5.0%	160 000 倍
β池	5 项贷款抵押中 4 项违约为输，其他情况为赢	0.003%	5.0%	1.684 倍

池 γ	5 项抵押贷款中 3 项违约为输， 其他情况为赢	0.1%	5.0%	44 倍
池 Δ	5 项抵押贷款中 2 项违约为输， 其他情况为赢	2.1%	5.0%	2.3 倍
池 ε	5 项抵押贷款都不违约为赢， 其他情况为输	20.4%	5.0%	0.2 倍

上述两种假设哪一种更行之有效，要看经济发展的状况。如果经济和房地产市场运转良好，则第一种假设（即 5 项贷款之间各自独立、互不相干）更合理。之所以会不时地发生违约情况，只是因为运气不佳：有人花了一大笔医疗费，或者有人失业了。但某一个人的违约行为与其他人的违约行为没有多少关系。

假设情况并非如此，而是存在某种共同的因素将这些房屋所有者的命运紧紧地联系在一起。例如，严重的房地产泡沫导致房价暴涨 80%，基本住房条件却没有任何的实际改善。现在你的麻烦来了：如果一个借款人违约了，剩下的几位借款人很可能也会出现同样的问题，那么你输掉赌局的可能性就会渐次提高。

美国在 2007 年年初经历了第二种假设中提到的状况（本章后半部分将会对房地产泡沫进行一个简短分析），但评级机构却将赌注押在了第一种假设中提到的那些基本上没有什么关联的风险上。尽管在房地产泡沫破灭之前的很长一段时间里，很多专家、学者已经意识到第一种假设的缺陷，有些评级公司的内部人士也对这些缺陷进行了揭秘，但那些评级公司却没有对这些缺陷做出充分解释。

举个例子来说，穆迪国际曾专门拿出一段时间对其预测模式进行临时调整，将 3A 级证券的潜在违约率提高了 50%。这一改变似乎十分谨慎，但提高的 50%真能弥补这一假设本身的漏洞吗？

若评级机构预测中发生的错误本身是线性的、可运算的，事情也许就好办得多了。但举债经营（即贷款投资）常常会使预测中的错误的复杂性增加很多倍，也存在引发呈几何倍数增长的非线性错误的可能性。穆迪公司对违约率所做出的 50%的调整，就好比是只涂了层防晒霜就告诉大家这样可以防核辐射一样，根本不足以应对这一严重的问题。这不仅意味着评级机构对违约风险的估计可能比实际违约率低很多，误差远不止 50%，而且极有可能是 500%乃至 5 000%。实际上，抵押贷款的违

约率要比评级公司宣称的高出近 200 倍，这就意味着它们的预测模式偏离实际数据的概率高达 20 000%。

从更广泛的意义上讲，评级公司的问题在于，它们无法区分风险和不确定性的不同，或者它们对两者间的差别根本就不关心。

“风险”一词由经济学家弗兰克·H·奈特于 1921 年第一次明确提出，我们可以为其定价。举个例子来说，玩德州扑克时，若对方“顺子”中缺一张牌，你就赢了，出现这种情况的精确概率为 1/11，这就叫风险。打牌时碰到邪门的“愤输牌”会很不愉快，但至少知道这种冤枉牌出现的概率，可以提前想好应对策略。从长远来看，当你的对手对差牌出现的概率的估计严重不足时，你就可以大赢特赢了。

不确定性是指难以度量的风险。也许你能隐约感受到暗藏的危险，甚至有时对这种危险的感觉很强烈，但究竟危险有多少，什么时候会爆发，这些问题你都不确定。粗略估计的误差可能是 100%，也可能是 1 000%。真实值究竟是多少，我们无从而知。这就是不确定性。风险是自由市场经济发展的助力，而不确定性则是阻力。

评级公司有这样一种魔力，能将捉摸不定的不确定性转变成看似可以感知的风险。那些刚刚发行的证券会受到系统不确定性的巨大影响，而评级公司却选择评估这些证券，并宣称它们可以对其风险进行量化。不仅如此，评级机构做出的所有评估结果几乎都令人瞠目：这些证券无投资风险。

有太多投资者将这些言之凿凿的结论误当作准确的论断了，而极少有人为意外情况未雨绸缪。

然而，尽管那些评级公司确实需要对金融危机承担很大的责任，但它们绝对不是酿成这些错误的唯一罪魁祸首。整场金融危机的悲剧可以分为 3 幕。

第一幕：房地产泡沫

在美国，房地产的投资获利历来就算不上丰厚。事实上，根据罗伯特·希勒及其同事卡尔·凯斯建立的凯斯–希勒函数来看，长期以来，美国房产的市场价格几乎没有任何上扬。因为通货膨胀，1896 年价值 10 000 美元的房产到 1996 年的价值仅为 10 600 美元。长达一个世纪的房地产投资，其收益率仅相当于股市一年的收益率。

尽管房地产投资的收益不高，但至少还算是一项比较稳妥的投资。进入 21 世纪前后，美国房价经历了史上最大的一次上涨，房产的涨价幅度可与第二次世界大战期间的房价增速相媲美，以 1942 年的房价最低点为参照点，当时的涨幅高达 60%。



图 **1–2** 凯斯－希勒函数，**1890~2006** 年美国房价

20 世纪 50 年代的房地产繁荣与 21 世纪初的房地产泡沫几乎没有任何共同之处，对比来看，我们就会明白为何 21 世纪初的房地产业如此混乱。

第二次世界大战后的几年，美国人的居住格局发生了巨大变化。美国人的生活从第二次世界大战前的过度储蓄转变为战后的消费激增，人们纷纷要求拓展居住空间。1940~1960 年这 20 年间，美国的自有住宅率从 44%猛增至 62%，主要集中在城郊地区。房地产繁荣的同时，也迎来了婴儿潮：战后，美国人口以每 10 年 20%的速度增长,这个增长率是 21 世纪初人口增速的两倍。这便意味着当时美国的房屋业主数量在 10 年内增加了 80%，这个增幅与房价的增速相吻合甚至超过了房价的增速。

相比之下，21 世纪初的几年，美国的自有房屋率增速缓慢，增幅最大的是 2005 年，这个增长率也只是从 10 年前的 65%增加到 69%，增幅不过 4%而已。从前没能力买房子的那些美国人当时还是没有几个能买得起房的。2000~2006 年的 6 年间，有 40%的房地产投资的收益名义上增长了 15%，但这个增幅连通货膨胀都抵消不了，更不用说再买一套新房子了。

然而，通过投机商炒房，以及有史以来最可疑的贷款和信誉度最差的消费者，美国房地产市场被人为地制造出虚假繁荣的景象。21 世纪初的几年内，全美存款率降至历史最低点，某些年份甚至低到只有 1%多一点。但是，得到抵押贷款却比从前任何时候都容易。房价不再受供需关系的制约，因为借贷者、股民以及评级机构等所有能从房屋买卖中获利的人，都在努力维持着繁荣的假象。

美国从未有过这样的房地产泡沫，但其他国家出现过这样的情况，当然结果都是同样惨烈。希勒对荷兰与挪威等几个国家几百年来的数据进行了研究，结果发现，一旦房地产行业的发展超出人们的负担能力，房地产市场必将崩盘。例如，20 世纪 90 年代日本臭名昭著的房地产泡沫，就是美国新近出现的房地产市场泡沫梦魇般的前车之鉴。1981~1991 年短短 10 年时间里，日本商业地产的价格飙升了 76%，但接下来的 5 年里又骤降 31%，这与美国房地产泡沫产生前后的房价走势如出一辙（图 1–3）。

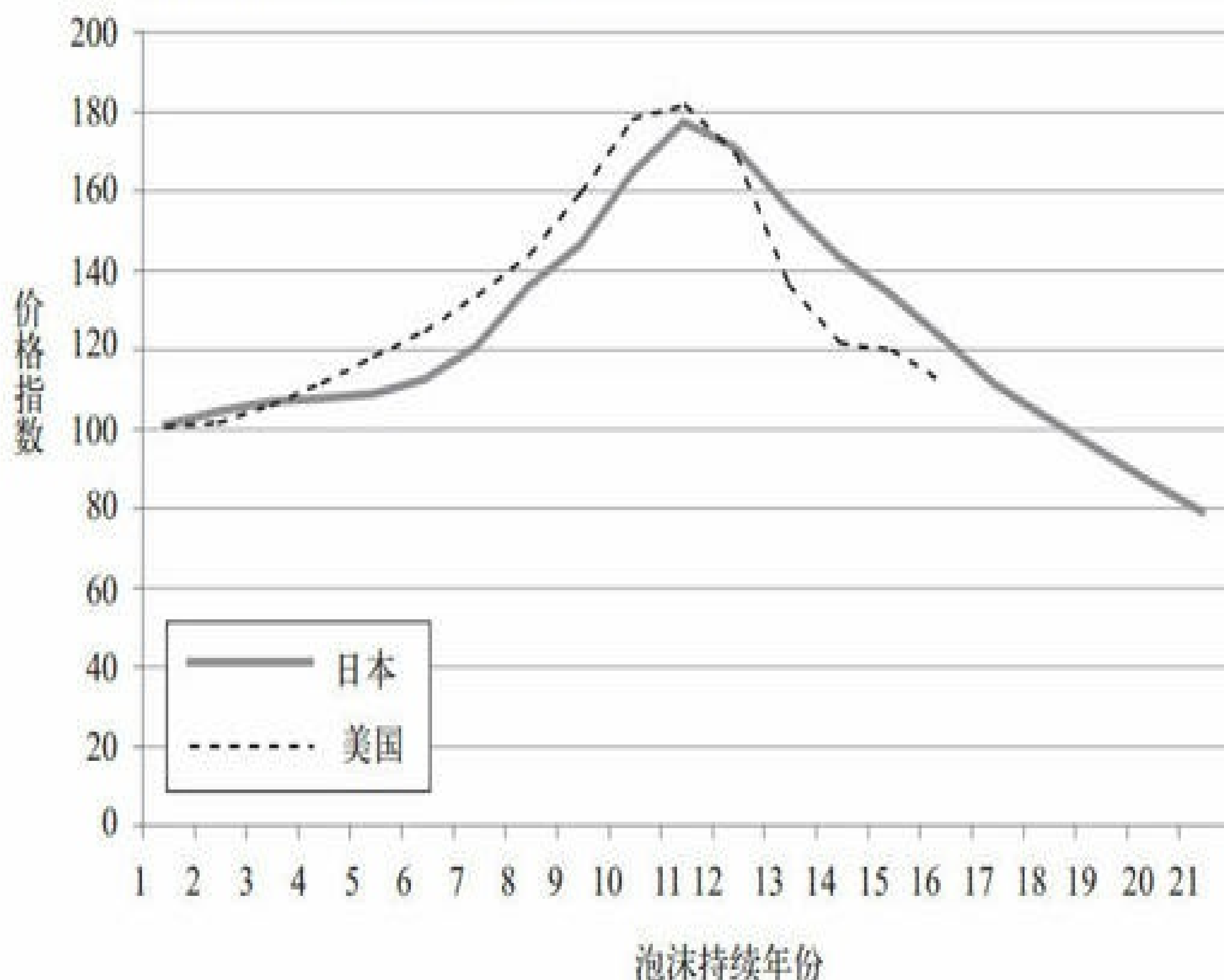


图 **1–3** 日本商业地产泡沫（**1981~2001** 年）和美国房价泡沫（**1996~2011** 年）

希勒还发现另一个导致房地产泡沫的重要证据：民众购买房屋时对这项投资可能带来的回报，总会怀着不切实际的期望。2003 年，凯斯和希勒负责的一项调查发现，许多业主希望他们的房产每年能增值13%。实际上，将通货膨胀因素考虑在内，1896~1996 年这 100 年内房价只增长了 6%，每年房价的增长率不过区区 0.06%。

这些业主对房市的过分自信也许是情有可原的，因为房地产泡沫已经渗透文化层面。2005 年，美国在 10 天内就先后开播了两档家装房地产真人秀电视节目：一档名为“玩转此屋”，另一档名为“玩转彼屋”。连那些本来无意从房地产投资中获取巨大利益的房屋购买者也开始行动起来。长期与希勒共事的经济学者乔治·阿克尔洛夫（2001 年诺贝尔经济学奖得主）的办公室位于加利福尼亚大学伯克利分校，那里的房价跌得最惨。阿克尔洛夫告诉我说：“我记得 20 年前，去萨克拉曼多一路通

畅，但现在常常交通堵塞，于是人们就想，如果我现在还不出手买房，5 年后同样的价钱就只能买离公路 16 公里远的房子了。”

不管那些房主认为他们买房不可能赔本还是不可以延迟购买，房地产市场的风险依旧那么自顾自地逐月提升。2007 年年末，问题浮出了水面：美国综合实力排位前 20 位的房地产公司中有 17 家房价当年出现下跌情况。更糟的是，反映房地产需求的主要指标——住宅营建批准数急剧下降，比高峰时减少了 50%。与此同时，债权人终于看清次贷市场标准松懈带来的后果，不愿再放贷。到 2007 年年底，全美房产止赎率翻了一番。

决策制定者的第一反应却是让泡沫再膨胀。佛罗里达州是受冲击最严重的州之一，但州长查利·克里斯特仍计划给新购房者发放 10 000 美元的贷款。2008 年 2 月，美国国会通过一项法案，希望通过大幅度增强房利美和房地美的放贷能力来刺激住房销量。但事与愿违，房价下跌的趋势无法逆转，2008 年住房销量下跌 20%。

第二幕：负债经营

尽管房地产泡沫刚产生时，就有很多经济学家意识到这个问题，但能够预见到房价崩盘会影响到整体经济的人为数不多。2007 年 12 月，《华尔街日报》预测版面的几位经济学家预测，2008 年经济衰退的可能性只有 38%。这一预测之所以令人印象深刻，是因为随后的数据显示，当时经济已经开始衰退了。而专业预测调查版面的那些经济学家更离谱，他们认为经济危机发生的概率还不到 500 分之一，但实际上早已危机深重了。

上述经济学家之所以做出了错误预测，是因为他们漏掉了两个主要因素。第一是房价下跌对普通美国民众资产的影响。截至 2007 年，美国的中产阶层家庭的财产中有超过 65%都花在自家的房子上了。然而，他们的生活却每况愈下——他们一直把房产权视为自动取款机。2001~2007 年这 6 年的时间，美国中产阶层家庭的非房产资产，即存款、股票、养老金、现金以及其他小生意的收入总和下降了 14%。所以，当房地产泡沫破裂使他们丧失了房产权时，这些中产阶层家庭发现，他们的经济状况与几年前相比可以说是极度恶化。

消费者更加现实的理财观使得消费性支出大幅下滑——经济学家称之为“财富效应”——下滑幅度大概相当于年国内生产总值的 1.5%~3.5%，这种状况很有可能使经济从慢速增长变为衰退。但是，普通的衰退是一回事，全球性金融危机则是另外一回事，两者不可同日而语。“财富效应”并不足以解释房地产泡沫是如何触发这样一场危机的。

事实上，房地产市场只是金融体系中相当小的一部分。2007 年，美国房屋销售总额约为 1.7 万亿美元，与股市每年创造的 40 万亿美元的交易额相比，可谓“小巫见大巫”。但与普通百姓的保守投资形成鲜明对照的是，华尔街金融大亨在房地产市场中的投资活动却是如火如荼。2007 年，住房抵押贷款支持证券交易总额约达 80 万亿美元，这就意味着只要有人自愿地将资金投放到抵押贷款中，每投入 1 美元，华尔街私下里就会押下价值 50 美元的赌注。

现在清楚金融危机的成因了吧：购房者的赌注被放大了 50 倍。这个问题可以概括为一个词：负债经营。

如果你从别人那里借来 20 美元，赌印第安人能打败西部牛仔，这就是负债打赌。同样的，你借钱获得贷款，或你将借来的钱投在一只抵押贷款支持的证券，也算是一次负债打赌行为。

2007 年，雷曼兄弟的负债率约为 33∶1，即 1 美元的资本要执行 33 美元承担的财政职能。这就意味着，如果雷曼兄弟的证券投资组合的价值只下降 3%~4%，公司拥有的资产净值就会变成负数，濒临破产。

高负债经营的不只是雷曼兄弟一家，另外几家主要的美国银行的负债率也都为 30%，并且在金融危机突然爆发之前这个数字还在持续上升。尽管美国银行负债率的历史数据参差不齐，但英格兰银行对英国的各大银行进行过一项分析，这项分析表明，现在金融系统的整体负债程度要么达到了 2007 年史上最高值，要么就是一个前所未有的新高度。



图 **1–4** 房地产市场抵押贷款支持证券

然而，与其他银行相比，雷曼兄弟的明显不同之处在于，这家投资银行对抵押贷款支持证券的“胃口”太大了。2007 年，雷曼所持有的 850 亿美元抵押贷款支持证券已经是公司资本基础值的 4 倍多，这就意味着，只要证券价值下降 25%就足以让公司破产。

正常情况下，广大投资者最不愿购买这样的资产了，至少他们会很

谨慎地避免赌局失手。

乔治·阿克尔洛夫告诉我说："如果有人在市场上竭力向你推销一款你完全不了解的产品，你就该想到他们在向你推销柠檬（没用的东西或者质次价高的产品）。"

阿克尔洛夫曾就这个问题写过一篇名为"柠檬市场"的著名论文，并且凭借这篇文章获得了 2001 年度的诺贝尔经济学奖。那篇论文指出，信息不对称将导致市场瘟疫，商品质量会下降，市场将充斥着非法的卖家和急切且轻信的买家。

假设大街上有个陌生人走上前问你有没有兴趣购买他的二手汽车，他一味地向你展示产品目录中的标价，却不愿让你试驾。对此，你难道不心存疑虑吗?这种情况的核心问题就在于，卖家比买家更熟悉这辆车的车况——返修率、行车里程等。不管对方出价多少，明智的买家不会做这笔交易，因为此时的不确定性要大过风险。你知道对方会给你很大的折扣，但却不知道这个折扣究竟该有多大。对方出价越低，你就越怀疑这笔交易的可信度，觉得这么好的事情不可能是真的。这世上也许根本就没有什么合理的价值可言。

但是，现在假设那个向你推销汽车的人有一位担保人，担保人看上去诚实可靠、值得信赖——这人可能是你的至交好友，也可能是你之前的生意伙伴，在这种情况下你就很可能会考虑做这笔交易了。评级机构就扮演着这样的角色，它们给那些抵押贷款支持证券做了很多 3A 评级作为担保，帮助它们赢得了原本不存在的商机。债券市场本来期望评级机构是《周六夜现场》中的戴比·唐纳，能多带来点"负面"消息，但评级机构俨然成了受人欢迎的美国演员小罗伯特·唐尼，左右逢源。

尤其是对雷曼兄弟公司，评级机构本该任用专业的驾驶员指引道路。在 2007 年 3 月的一次电话会议中，雷曼兄弟公司的首席财务官克里斯托弗·欧米告诉投资者，他对近期市场的停滞并不担心，别人急于清算之时正是抄底大赚一笔的好时机。他解释道，贷款市场的信誉度是很可靠的，但这一结论只能从那些证券的 3A 评级中得出，而不是从作为次级附属担保物的次级抵押贷款的质量中体现出来的。这一次，雷曼兄弟公司算是买了一个烫手的"柠檬"。

一年之后，房地产泡沫开始破灭，雷曼公司迫切希望转手卖掉这些"柠檬"，但那些投资者要求用作信用违约掉期的保险费数额激增——一旦违约，这笔费用便会派上用场，也可以作为一种基本的保障手段，防止违约情况的发生——这些费用只能弥补雷曼兄弟公司 20%的资金缺口。所有努力都化作枉然，也为时过晚，2008 年 9 月 14 日，雷曼兄弟公司宣布破产。

幕间休息：从贪婪到恐惧

雷曼兄弟公司破产之后的种种真实经历，足可以写成一本书了（也确实被写进了《大而不倒》等优秀著作中）。这次事件也足以让人们谨记，一个金融公司倒下后，留下的大笔债务会一直困扰着经济发展。雷曼兄弟公司无力填补证券投资失败的漏洞，就需要由他人埋单，这又进一步波及了其他公司，从而对整个金融系统造成冲击。投资者和债券发行者眼看着危机发生，却不知道究竟谁欠了谁的钱，更无从分辨哪些公司还有偿还能力，于是，不管利息多高，他们都不愿意再借出一分钱，连经营状况良好的各大公司的运转也由此受到了影响。

正是出于这个原因，很多政府有时会拯救那些难以自保的金融公司，虽然这样做会使纳税人背上巨大的负担，还会损害政府自身的形象，但它们还是会出手。美联储就曾挽救过贝尔斯登公司和美国国际集团两大金融服务公司，然而这次却没有向雷曼兄弟伸出援手，广大投资者的期待都落了空，于是第二天道·琼斯指数一开市就暴跌500点。

为什么美国政府出手帮助贝尔斯登公司和美国国际集团，却对雷曼兄弟公司不闻不问呢？个中缘由我们无从得知。一种解释说，因为雷曼兄弟公司长期以来都不负责任，金融状况每况愈下，因此政府不确定要付出多大的代价才能挽救它，于是就不想做费力不讨好的事了。

2009 年 12 月，我在白宫遇见了时任美国国家经济委员会主席的劳伦斯·萨默斯，他告诉我说如果政府在危机发生时保住了雷曼兄弟公司，可能美国的状况就会好不少，但当时金融系统的负债经营率实在太高了，不得已只能忍痛割爱。

提到金融危机时，萨默斯说：“这是献身克己的预言。过去大家基本上都在负债经营，而一旦所有人都在负债经营，经济体系就必然是脆弱的，事实证明，他们的自满情绪根本就是毫无根据的。雷曼兄弟公司就是丢在干燥森林里的一根点燃的香烟，成了金融危机的导火索。即便倒下的不是雷曼兄弟，也很可能是别的公司。”

萨默斯认为美国经济由一系列反馈回路组成，供需关系就是其中简单的一环。假设你正在经营一家卖柠檬汽水的小店，若降低价格，销量就会增加；提高价格，销量就会下降。如果你赚得盆满钵满只是因为户外的温度快达到 38 摄氏度，而附近卖柠檬汽水的店铺只有你一家，这就有风险了。若某个恼人的家伙在街对面也开了一家汽水店卖柠檬水，你就得降价了。

供需关系的例子属于一种负反馈：价格走高，销量就会下降。尽管称为“负”反馈，这种反馈对市场经济来说却是积极有益的。假设情况相

反，供需关系总是正反馈，即价格提高时，销量也会提高。你将柠檬汽水的价格从 25 美分涨到 2.5 美元，销量非但没有下降，反倒增加了一倍。好了，现在你又涨价了，将柠檬汽水的价格从 2.5 美元涨到 25 美元，销量又翻倍了。最终结果就是，你将一杯柠檬汽水卖到了 46 000 美元——这是美国的人均年收入——3 亿美国人全都排大队来买你的汽水。

这是一个正反馈的例子。这种情况乍一看觉得很开心，但很快你就会发现全美国人民要倾尽所有才能买得起一杯柠檬汽水，你本想用卖柠檬汽水的钱去买自己一直十分钟情的电子游戏，可是却没有人愿意离开队伍为你生产这些游戏。

萨默斯认为，通常负反馈在美国经济中占据主导地位，如同恒温器一般防止经济衰退或过热。他还认为，恐惧与贪婪也是最重要的反馈之一。有些投资者谨慎小心，有些则敢于冒险，这样市场才能平衡：如果股价因为公司的财政状况恶化出现下跌，胆小的投资者就会纷纷抛售，而贪婪的股民则会乘机抄底。

然而，贪婪和恐惧是两个非常不稳定的因素，只有两者保持平衡，经济才能顺利发展。若贪婪在经济体系中占上风，就会产生经济泡沫；若恐惧因素压过贪婪，经济又会陷入恐慌。

一般来讲，我们作决定前咨询一下亲朋好友的意见是有益处的，但问题是，若他们的判断是无奈之选，那我们自己的判断肯定也好不到哪里去。人们评估自己房子的价格时常与他人的房子相比，如果一套位于城市新区的三居室房屋售价为 400 000 美元，那附近的一座殖民时期的房子若要价 350 000 美元，就会让人觉得天上突然掉了一个大“馅饼”，捡了大便宜。依此来看，如果一套房子的价格上涨了，就极有可能拉动周边其他房屋的价格。

或者假设你正在考虑投资另外一种资产，比如抵押贷款支持证券。这类商品也许更难估值，但购买这种资产的投资人越多（并且为其担保的评级公司越多），你就越有可能相信这些资产是安全的，是值得投资的。因此，你得到一个正反馈，但同时也有可能遭遇经济泡沫。

但是，最终主导房地产市场的还是负反馈：根据现有价格，没有哪个美国人能买得起房。这样看来，很多已经买了房子的美国人一开始根本就不可能买得起房子，但很快他们就靠抵押贷款买了房子。人们都在假设所有买进这些资产的人不可能出错，最终这些资产吸引了数十亿美元的贷款，但这种高负债的局面必然会对经济造成严重破坏。直至泡沫破灭时，人们才如梦方醒。

萨默斯曾在 2009 年告诉我说：“过去我们不算恐惧，但是太贪婪；

而今我们不算贪婪，但是太恐惧了。”

第三幕：这次还是犯了同样的错误

一旦房地产泡沫破灭，那些贪婪的投资者将会发现到处都潜藏着不确定性，他们的贪婪就会转为恐惧。摆脱金融危机的过程——所有人都竭力想弄清楚到底是谁欠了谁——会带来一些遗留问题，而且这些问题会持续相当长的时间。美国经济学家卡门·莱因哈特和肯尼斯·罗格夫为了撰写《这次不一样：800 年金融危机史》一书，研读了大量金融史资料，他们发现，数次金融危机导致的失业问题通常都会持续 4~6 年。莱因哈特还进行过另外一项研究，关注了距离现在比较近的几次金融危机，他发现，经历这些危机的 15 个国家中，有 10 个国家的就业率始终都无法恢复到危机前的水平。这样一来，金融危机与正常的经济衰退的区别便一目了然了。正常的经济衰退发生后，随着经济元气的不断恢复，通常在下一年，经济增长就会超过一般水平，就业率也会迅速赶超衰退前的水平。尽管将金融体系和经济体系的其他部分区分开来很重要，但是很多经济模式并未对此作过区分。

美国政府本该好好地从莱因哈特和罗格夫的书中吸取教训，然而他们并没有这样做。很快，美国政府就尝到了自己糟糕预测带来的苦果。

2009 年 1 月，巴拉克·奥巴马即将宣誓就职，由美国经济顾问委员会主席克莉丝汀·罗默和萨默斯领衔的新的白宫经济团队也开始就位。新团队奉命制定一系列的经济刺激计划，以刺激企业及个人的消费需求。罗默认为政府需要为此投入 1.2 万亿美元。但白宫政治团队提出异议，他们认为很难说服国会接受上万亿美元的价码，于是这个数字最终被修改为 8 000 亿美元。

为了向美国国会和普通民众推销这一计划，罗默和她的同事准备了一份备忘录，上面记录着危机的严重程度以及这份刺激计划将采取何种措施来缓解这次危机。这份备忘录的显著特点就是，其中配有很多预测图，明确地呈现出在有刺激计划和没有刺激计划这两种情况下的失业率走势。备忘录中写着，如果没有经济刺激计划，2008 年 12 月发布的 7.3%的全美失业率，到 2010 年年初将达到约 9%的最高点；但如果有了刺激计划，失业率就永远不会超过 8%，而且最快会在 2009 年 7 月开始降低。

2009 年 2 月，在按政党画线的投票中，美国国会通过了这一刺激计划，但是美国的失业率仍持续上升，到 2009 年 7 月失业率升至 9.5%，同年 10 月更是达到令人咋舌的 10.1%，这比白宫之前预测的“没有刺激计划”的后果还要严重。保守党的博主们每月都会对罗默的预测图进行

厚颜无耻的更新，当然是将实际失业率叠加在原来那些过于乐观的预测图上进行修改。

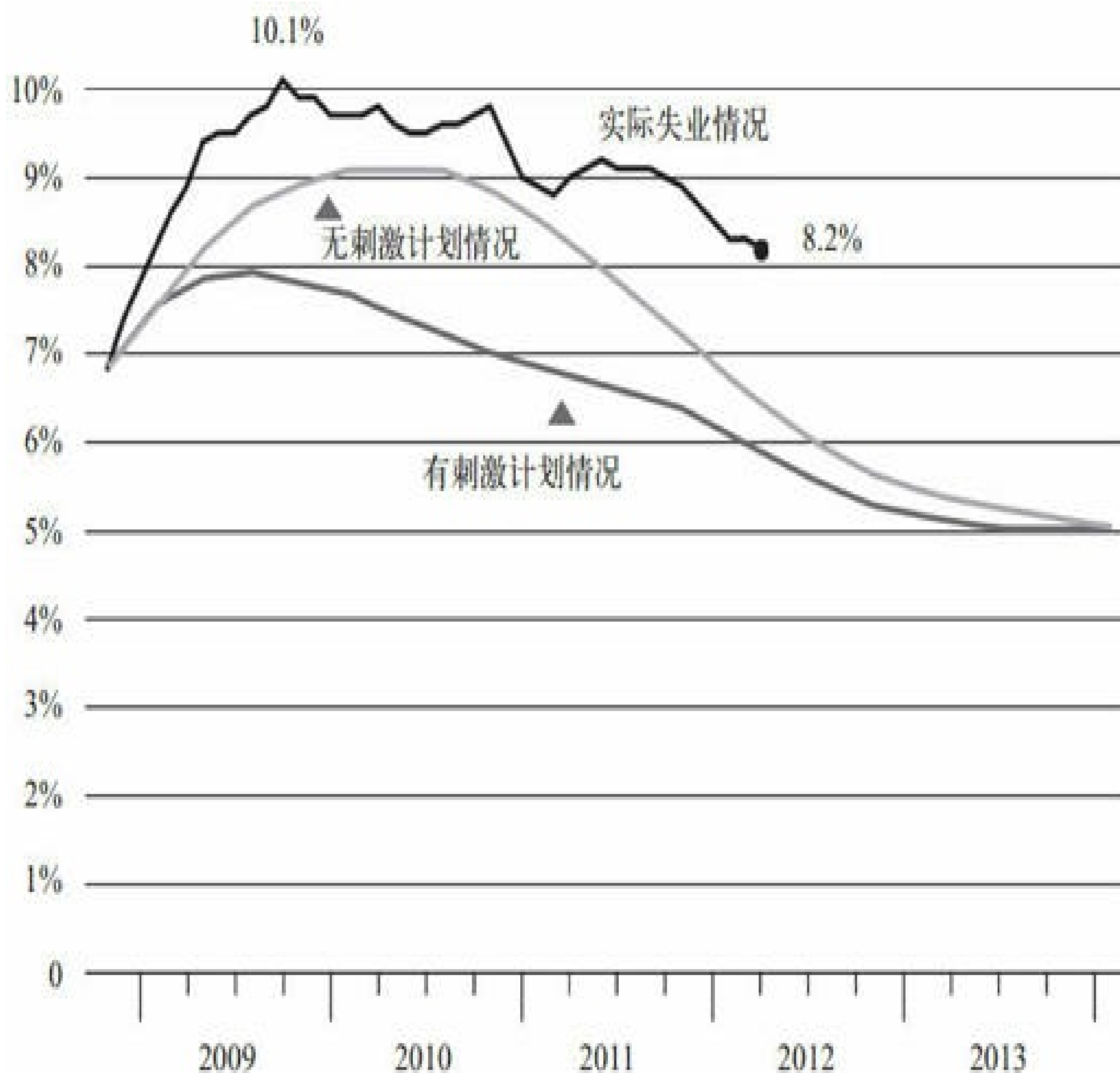


图 **1–5 2009** 年 **1** 月白宫制定的经济刺激计划

资料来源：（美国）劳动统计局

然而，普通民众看到这幅预测图后，得出的结论与罗默的说法有所不同，事实上，两者的结论根本就是截然相反的。但经济学家保罗·克鲁格曼一开始就认为这项刺激计划的投资明显不足，他认为此次刺激计划的失败证明，白宫过分低估了个人及企业需求的降低幅度。克鲁格曼说："尽管采取了特别的经济刺激手段，失业率却并没有明显降低，这一事实让我们明白，我们面对的是金融危机的致命冲击。"当然，其他

经济学家都没有看到这一点，他们认为这些图只是证明刺激计划彻底失败了。

白宫可以像标准普尔评级机构一样为自己辩护：“出错的不单是我们，别人也都犯了同样的错。”的确，白宫的预测与当时很多独立经济学家发布的预测结果如出一辙。另外，最初的经济统计数据明显严重低估了此次危机的严重程度。在实施经济刺激计划期间，罗默和萨默斯得到的第一份政府评估结果显示，2007 年年底，美国国内生产总值的下降率为 3.8%。但事实上，此次金融危机对经济造成的危害比这个评估结果多出一倍以上。美国国内生产总值的实际下降率曾一度接近 9%，这就意味着这个国家的国内生产总值比政府的第一份评估结果实际上要少 2 000 亿美元。

也许白宫犯下的更不可宽恕的错误是，他们做出这样一个看似很精确的预测，却没有告诉人们这个预测也有可能会出错。白宫内外的经济学家都没能成功地预测失业率等主要经济指标的走势。（在本书第六章中，我将详细论述宏观经济预测的问题）。经济衰退时所作的失业率预测，因为不确定性而常常存在 2%的误差。所以，即便白宫认为 8%的失业率是最有可能出现的结果，这个数字也极有可能攀升至两位数（或也有可能会降至 6%的低水平）。

为刺激消费所投入的资金的效力也具有很大的不确定性。不同研究得到的乘数效应——为刺激消费而投入的每一美元对经济增长的贡献——的预估结果也有出入。一些研究认为，投入一美元会带来高达 4 美元的国内生产总值的增长，另一些研究则认为，一美元只能带来 60 美分的回报。任何形式的宏观经济预测本身都具有很大的不确定性，一旦以这种不确定的预测结果为基础对刺激手段的有效性进行评估，就相当于把两种很大的不确定性叠加在一起，此时的预测就很有可能失败。

失败的预测都是非样本预测

此次金融危机的预测至少存在 4 大失败之处：

- 人们本来可以发觉这次的房地产泡沫是一个错误的预测，但那些房主和投资者却错误地认为，不断走高的房屋价格表明房屋价值还会不断提高。事实上，历史证明，房屋价格走高时，其价值反而更容易降低。

- 各大评级机构以及像雷曼兄弟这样的投资银行没有预测到抵押贷款支持证券的巨大风险。但问题是，这些评级机构之所以预测

错误，并不是（他们在美国国会听证会上辩解的那样）由于他们没有发现房地产泡沫的存在，而是由于其预测模式存在着很多错误的假设，对这次房价暴跌可能带来的风险估计不足、盲目乐观。

· 人们普遍没有预料到一次房地产危机竟会触发一场全球性的金融危机。这次危机是由市场中的高负债经营导致的，只有在 1 美元的风险有 50 美元的风险资金支持的情况下，美国人才愿意将资金投到一所新房子里。

· 金融危机发生后，人们没能预测到这次危机可能引发的巨大经济问题。莱因哈特和罗格夫已经发现金融危机常会导致严重而长久的经济衰退，但许多经济学家和政策制定者对这个观点并不在意。

这些失败的预测有一条共同线索，即在上述每种情况下，人们在评估数据时都忽视了一个重要因素：

· 近些年来，美国的房价从未出现过大幅回落，这种情况使得很多房主对房价信心满满。然而，美国的房价在房市崩盘前高得离谱，这种提升幅度是前所未有的。

· 穆迪和标准普尔等评级机构过去总能对各种金融资产做出准确的测评，基于此，各大银行对这些评级机构对抵押贷款支持证券的评估能力毫不怀疑。然而，这些评级机构此前从未为这种复杂的新型信用违约率证券做过测评。

· 过去，房价的浮动并未对金融体系造成多么严重的影响，于是许多经济学家便对金融体系承受房地产市场危机的能力充满信心。然而，金融体系或许从未出现过这样高负债经营的局面，自然也就从来没有为房地产市场投入过这么多的资金。

· 最近几次经济衰退几乎都呈现出“V”字形的迅速恢复态势，这种经历使那些政策制定者坚信，此次金融危机之后，经济也会迅速恢复。然而，之前的那些经济衰退都和金融危机无关，其严重程度不可与这一次同日而语。

此类问题有一个专门术语：以上这些预测者所考虑的事件都属于“非样本”预测。一个预测一旦出现严重错误，其中通常都会存在“非样本”预测这一问题的痕迹。

“非样本”究竟是什么意思呢？我们举个简单的例子解释一下。

失败预测的公式——非样本，无思考

假设你是一名非常出色的司机。每个人都认为自己是一个好司机，但是你有实际的驾驶记录可以证明这一点：驾龄长达 30 年，在 20 000 次出行过程中，只发生过两次轻微剐蹭事故。

你也不是酒鬼，醉驾这种事情似乎永远不会发生在你的身上。但是，有一年在公司的圣诞派对上，一位好友要离开公司，你当时的工作压力又很大，于是情绪出现了波动，不知不觉间喝了 12 杯伏特加，你喝醉了。此时该开车回家，还是叫一辆出租车呢？

这个问题的答案肯定是叫辆出租车载你回家。

但你突发奇想要自己开车回去，你是这样为自己找理由的：自己曾经有 20 000 次驾车出行的经历，只发生过两起小意外，其他 19 998 次都安全抵达目的地。安全率这么高，又何必那么麻烦让出租车载你回家呢？

但问题是，20 000 次出行记录没有一次是像这次醉酒驾车的，你的醉驾样本数量不是 20 000 次，而是零次。因此，用先前的经验预测此次驾车的风险是毫无根据的。这个例子就解释了什么是“非样本”问题。

这一问题看似很容易避免，但评级公司正是犯了这一错误。穆迪公司根据过去的数据——特别是美国自 20 世纪 80 年代以来的房地产市场数据——构建了一个预测模式，以此来评估不同抵押违约行为之间的关联程度。问题是，从 20 世纪 80 年代到 21 世纪初期的二十几年的时间里，美国的房价一直保持稳定或略有增长。在这种情况下，认为一个房主的抵押贷款与其他贷款没有什么关系，这种假设看起来好像没有什么问题。但这些从前的数据无法显示当房价开始不断走低时会出现什么情况。房市崩盘是“非样本”事件，评级公司在此基础上对违约风险进行评价的预测模式也就毫无价值可言。

前事不忘，后事之师

然而，穆迪公司当时并未竭尽所能，如果它们当时能放宽眼界，还是可以得到更合理的评估结果的。美国此前确实没有经历过房地产市场的崩盘，但其他国家经历过，而且结果惨不忍睹。如果穆迪公司对日本房地产泡沫后的违约率有所了解，就会更加理性地看待抵押贷款支持证券的危险性，也就不会给出 3A 评级了。

很多预测者通常都不愿意考虑“非样本”中存在的这些问题。当我们将样本扩大到涵盖了那些在时空上都离我们很远的事件时，这通常意味着我们会遇到一些自己并不熟悉或与所进行研究关系并不紧密的案例。因此，这一模式看上去就不那么有说服力，展示在幻灯片（或期刊论文

或博客）中时，也不会给人留下很深刻的印象。在这种情形下，我们只好承认自己对这个世界的了解并没有想象的多，而我们的个人感受和职业动机几乎总在阻止我们这样做。

我们忘了——也可能是故意忽略——我们的预测模式就是简化了的世界，我们以为即使犯了错，也无大碍。

然而，在复杂的系统中，错误都不是用程度来衡量的，而是用级别衡量的。标准普尔和穆迪将债务担保凭证的违约风险低估了 200 倍，那些经济学家认为发生概率只有 500 分之一的经济衰退也实实在在地上演了。

就像我在引言中提到的那样，在信息时代我们面临的一个重大挑战，就是全球的知识总量在增加，而我们实际掌握的知识和自认为掌握的知识之间的鸿沟却越来越宽。这一症状与那些貌似精确实则错误的预测是有关联的。穆迪将计算结果保留到小数点后两位，看似精确却脱离了实际。这就好比你说自己是一位射击高手，因为你的子弹每次都能射中同一个位置，即使每一枪都离目标十万八千里，你也会这样讲。

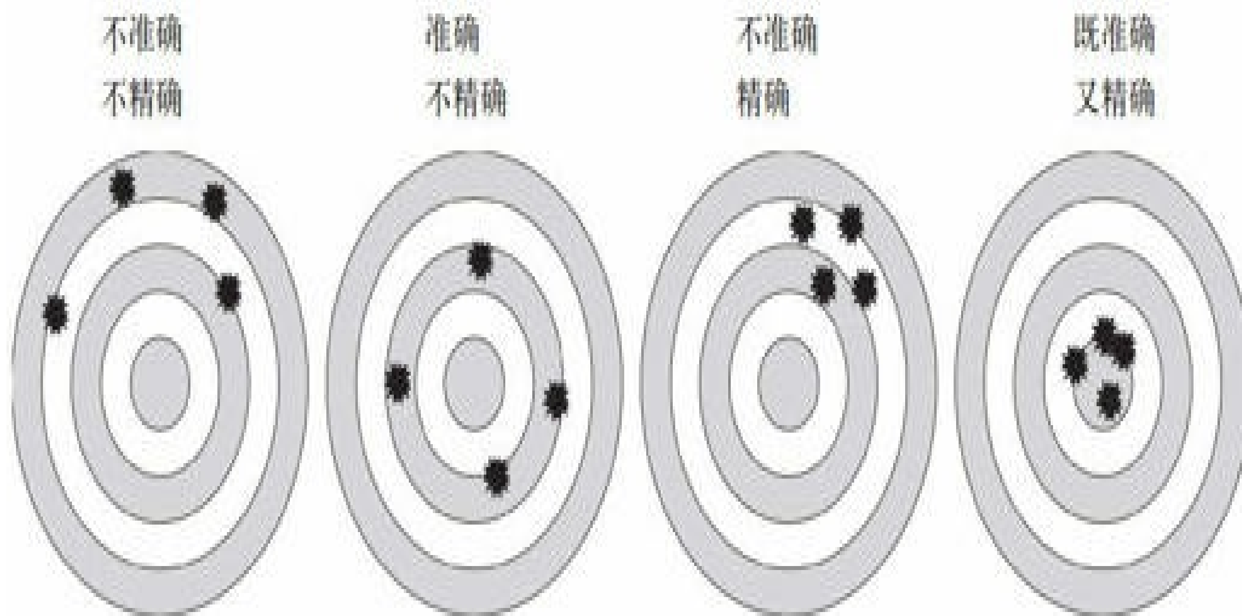


图 **1–6** 准确度与精确度对比

金融危机以及其他大多数失败的预测都源于一种盲目的自信。他们用精确的预测冒充准确的预测，于是我们有些人就上当了，还将赌注翻番。我们自以为已经克服了辨别力中的盲点，却没有料到实力雄厚如美国这样的国家此时经济的发展也戛然而止了。

第二章 政治选举预测：狐狸和刺猬，谁更聪明？

对许多人来说，政治预测就是“麦克劳夫伦讨论小组”这档电视节目的同义词。从 1982 年开始，每周日电视台都会播出这样一场“政治圆桌会议”，之后还会被“周六夜现场”节目恶搞一番。节目主持人约翰·麦克劳夫伦是一个 80 多岁的暴脾气老人，曾在 1970 年竞选美国参议员时落败。该节目像组织体育活动一样组织政治专家展开讨论，小组成员要在半个小时内循环讨论四五个话题，麦克劳夫伦会就相关话题对这些成员步步紧逼，这些话题从澳大利亚政治探讨到地外智慧生物探寻，无所不包。

“麦克劳夫伦讨论小组”每期节目的最后都设有一个“预测”的环节，在这个环节中，小组成员用几秒钟的时间讨论当天发生的某件事的重要性。有时，他们可以选择一个和政治关系不大的话题进行预测；有时，麦克劳夫伦会就某一个特定问题突击询问小组成员，让他们现场作答，麦克劳夫伦称此为“强迫性预测”。

麦克劳夫伦的有些问题很难回答，比如让小组成员从几个最佳候选人中选出最高法院的下届法官。其他问题则比较温和，比如 2008 年美国总统大选前夕，他曾让小组成员预测麦凯恩和奥巴马谁会当选。

答案显而易见。2008 年 9 月 15 日，雷曼兄弟宣布破产，引发了自 20 世纪 30 年代大萧条时期之后最严重的经济衰退，正是从那天起，奥巴马的选票数几乎在所有的美国投票站都领先麦凯恩。而且，不论是在俄亥俄州、佛罗里达州、宾夕法尼亚州或是新罕布什尔州这样的“摇摆”州，还是在一些民主党很少获胜的地区，如科罗拉多州和弗吉尼亚州，奥巴马也都处于领先地位。我为自己的网站开发的预测模式和其他类似的统计模式都表明，奥巴马赢得大选的可能性超过 95%，赌博市场虽然对美国总统大选结果的估计稍显模棱两可，但对奥巴马获选的支持率预测也达到了 7∶1。

而“麦克劳夫伦讨论小组”的首席成员帕特·布坎南对这个问题却避而不谈。他说：“一切自有定数，本周末自然见分晓”，这句话引得其他成员一阵爆笑。来自《芝加哥论坛报》的克拉伦斯·佩吉预测说：“这场大选难分伯仲。”《福克斯新闻》的莫妮卡·克劳莉更加夸张，她大胆预测麦凯恩会以 0.5 个点险胜。只有《新闻周刊》的埃莉诺·克里夫特点明了显而易见的事实，她预测奥巴马最终会当选。

接下来的那个周二，结果出来了，奥巴马以 365 张选举人票对麦凯恩的 173 张选举人票当选为下一届美国总统，这一结果与民意测验和统计模型的预测结果基本吻合。虽然不是以压倒性的优势获胜，但也绝不是“难分伯仲”：奥巴马以总选票数高出近 1 000 万张的实力击败麦凯恩，那些预测失败的人该想想怎样自圆其说了。

一周后，“麦克劳夫伦讨论小组”再次迎来上周那 4 位小组成员，这几位此次对上次预测失败的事情缄口不谈。他们讨论了奥巴马获胜的细枝末节，讨论了奥巴马任命拉姆·伊曼纽尔为下一届美国国务卿，还讨论了他与俄罗斯总统德米特里·梅德韦杰夫的交情。大家对上次失败的预测只字未提——尽管那次预测基本上与当时所有的证据背道而驰，并且还在美国国家电视节目中播出。事实上，小组的几位成员似乎自始至终都认为这样的结果是不可避免的。克劳莉是这样解释的：“2008 年本就是个换届年，麦凯恩的竞选宣传太不到位了。”克劳莉似乎忘了仅仅一周前，她还把赌注押在麦凯恩一边。

我们不能单凭一次预测就对一位预言家做出评判，但有一种情况属于例外。截止到美国总统大选前的那个周末，认为麦凯恩仍将获胜的假设中，唯一说得通的也许就是大多数人对奥巴马都存在种族敌意，而这一点在民意调查中无法体现出来。然而，所有小组成员都没有提到这个假设。他们似乎是在另一个世界进行预测，那里没有民意测验，也没有经济崩溃，在那里美国总统小布什仍然很受欢迎，完全没有连累到麦凯恩。

虽然如此，我还是决定研究看看这种情况是否正常。“麦克劳夫伦讨论小组”花钱请来那些以谈论政治为生的专家们，他们是否真有预测的本事？

我对这个节目最后的预测环节中的近千份预测报告进行了评估，这些预测都是由麦克劳夫伦及其节目的其他专家做出的。其中近 1/4 的预测，要么模棱两可，要么不切实际，想要分析都无从下手。我将剩下的那些预测从完全错误到完全正确分为 5 个级别。

结果，就好像那些小组成员作预测时用了投掷硬币的方法一样，完全错误或基本错误的预测共有 338 份，而完全正确或基本正确的预测也是 338 份。

表 **2–1**“麦克劳夫伦讨论小组”预测分析

完全正确	285	39%

大部分正确	53	7%
部分正确，部分错误	57	8%
大部分错误	70	10%
完全错误	268	37%
已评估预测总数	733	100%
未评估预测	249	

包括成功预测出 2008 年美国总统大选结果的克里夫特在内，小组成员的预测能力不相上下。我为每个成员计算了一个百分比得分，基本反映了他们预测成功的次数。克里夫特和其他 3 位参与该节目次数最多的小组成员——布坎南、已故的托尼·布兰科里，还有麦克劳夫伦本人——的分数几乎相同，预测成功比率都在 49%~52%。这说明他们预测成功和预测失败的概率是一样的，他们所表现出的政治敏锐度和从理发店走出来的四重唱演员差不多。

当然，从某种程度上来讲，“麦克劳夫伦讨论小组”这档节目不过是供政治狂热者消遣的低俗闹剧，当时此类“声高即有理”的节目很流行，以美国有线电视新闻网的“唇枪舌剑”节目为例，节目中自由派与保守派只是没完没了地与对方争辩。但是，录影棚时代与过去那个“声高即有理”的时代也没有什么不同，只不过现在自由派与保守派都在各自的频道中活动，在电视节目时间表中被美食频道和高尔夫频道隔离在和平地带。这样的安排虽然未见得会催生更加可靠的分析，但看似可以带来更高的收视率。

但是，栏目组花钱请来那些学者讨论问题，初衷是因为他们的判断更准确，学识更渊博，而不是因为他们发表观点时的嗓门够大。那些政治科学家和美国政府的智囊团中的那些分析师怎么样呢？他们是不是更擅长预测呢？

政治学家是名副其实还是徒有虚名?

苏联解体与东欧剧变的发生极为突然，但综合各方面的因素进行考虑，这些事情的发生又是顺理成章的。

1987 年 6 月 12 日，美国总统里根站在勃兰登堡门下，恳请苏联领导人戈尔巴乔夫拆除柏林墙，他的这一令人欢欣鼓舞的建言和当初肯尼迪承诺把人送上月球的誓言一样大胆。里根当时是有先见之明的，因为此后不到两年柏林墙就被拆了。

1988 年 11 月 16 日，爱沙尼亚共和国（面积同美国缅因州相仿）议会宣布脱离苏联独立。此后不到 3 年的时间里，戈尔巴乔夫否决了莫斯科强硬派主张的政变，苏联国旗最后一次在克里姆林宫前降下。爱沙尼亚和其他苏联加盟共和国不久后都脱离苏联，各自独立。

如果说苏联的衰亡是有迹可循的，是可以预测到的，然而几乎所有主流的政治学家都没有察觉到这一点。即使有，当时也会被当成笑料。如果那些政治学家连苏联解体都预见不到——也许是 20 世纪后半叶最重要的事件了——那他们还能做些什么呢？

当时在加利福尼亚大学伯克利分校任教的心理政治学教授菲利普·泰特罗克也有着同样的疑问。苏联分崩瓦解之前，泰特罗克就已经野心勃勃地展开了一项空前的实验。从 1987 年起，泰特罗克就开始从学术界和政府的各方面专家那里广泛搜集各类关于国内政治、经济和国际关系的预言。

泰特罗克发现，那些政治专家很难预测到苏联解体，因为既要预测到政权的衰亡，又要找到其衰亡的原因，这就需要进行预测的人将不同立场的观点论据穿插在一起。这些观点本身没有什么内在矛盾，但通常是由身处两个不同政治阵营的人发出的，而坚定地站在某一个思想阵营的学者则不可能同时接受两种思想。

一方面，戈尔巴乔夫很明显是这一事件的主角——他是真心诚意地主张改革的。如果戈尔巴乔夫当初选择做会计或是诗人而不是选择踏入政坛，苏联政权也许至少几年内还不至于垮台。自由派对戈尔巴乔夫还是心存同情的，可保守派不愿相信戈尔巴乔夫，有些人认为戈尔巴乔夫的公开讲话不过是故作姿态而已。

另一方面，保守派对共产主义有一种本能的批判。他们很快就认定苏联经济正在走下坡路，普通民众的生活正变得愈加艰难。1990 年，美国中央情报局估计——相当不准确的估计——苏联的国民生产总值约为美国的 1/2（按人均水平计算，与今天的韩国和葡萄牙这类稳定的民主国家的水平相当）。实际上，最新的证据表明，由于长期陷于阿富汗战争的泥潭，且中央政府对各类社会问题疏于管理，苏联的国内生产总值比美国中央情报局提供的数据还要低约 1 万亿美元，而且每年国内生产总值缩减的比例都会达到 5%，通货膨胀率也高达两位数。

综合两方面因素考虑，苏联解体其实很容易预见。开放媒体和市场，赋予公民更大的民主权利，戈尔巴乔夫为苏联人民提供了一种新机制，以促进社会制度的改革。由于国家经济百废待兴，人们自然乐于支持戈尔巴乔夫提出的机制。然而当时的中央政权已经不堪重负，根本承受不了这种改革：爱沙尼亚人厌恶苏联人，苏联人同样厌恶爱沙尼亚

人。各个加盟共和国为苏联经济贡献的力量远远低于他们从莫斯科得到的补助金。捷克斯洛伐克、波兰、罗马尼亚、保加利亚、匈牙利和民主德国都在 1989 年的改革之列，东欧这些国家一旦像多米诺骨牌一样一个接一个地倒下，不论是戈尔巴乔夫还是别的人，都无法阻止整个国家的土崩瓦解。许多苏联学者看到了这一问题的几个部分，但鲜有专家能把这些碎片拼接起来，因此，几乎没有人能预见到苏联会突然垮台。

受苏联解体的案例启发，泰特罗克开始到其他领域的专家那里进行调查，让那些专家做出各种预测，如海湾战争、日本房地产泡沫、魁北克脱离加拿大统治的可能性，几乎将 20 世纪八九十年代所有大事件都包括在内。对苏联解体的失败预测究竟是个特例，还是那些“专业”的政治分析者只是徒有虚名？泰特罗克的研究持续了 15 年之久，他最终将这些研究发表在 2005 年出版的《专家的政治判断力》一书中。

泰特罗克的结论招来了强烈谴责。他的调查中涉及的那些专家——无论职业、阅历或者研究领域——所作的各项预测的准确率跟碰运气差不多，对未来的政治事件进行预测时，他们预测的准确程度甚至不及那些尚不成熟的统计方法计算的准确度。他们过于自信了，计算概率的能力也很差：他们宣称不会发生的事件中有 15%实实在在地发生了，而他们认为绝对会出现的情况中则有 25%根本没有出现。他们是否在对国家的经济状况、国内政治或国际事务作预测这点并不重要，重要的是他们的预测全线溃败。

狐狸型专家：善于变通，更善于做出准确的预测

尽管那些专家的整体表现欠佳，但泰特罗克还是发现一些不错的专家。表现不佳的专家往往是被媒体引用预测言论最多的那些人。泰特罗克发现，如果一位专家接受新闻媒体的采访越多，他的预测就可能越不准确。

而另外一些专家的表现相对好些。曾经接受过心理学培训的泰特罗克对那些专家的认知风格很感兴趣，他想看看这些人是如何看待世界的。于是，泰特罗克从性格测试中选取了一些问题，让所有专家作答。

根据诸位专家对这些问题的回答，泰特罗克把他们分为两类，他称其为“刺猬”和“狐狸”。刺猬和狐狸的称呼参照了以赛亚·伯林所写的一篇有关俄国小说家列夫托尔斯泰的散文——《刺猬与狐狸》。而伯林则是借鉴了希腊诗人阿基罗库斯的作品才想到了这一题目：“狐狸千伎百俩而有尽，刺猬凭一技之长而无穷”。

除非你是托尔斯泰的崇拜者，或是对华丽的散文如痴如醉，否则你肯定不会读伯林的散文。不过基本思想是，作家和思想家总是分属两大

范畴：

刺猬属于A型性格的人，他们相信“凭一技之长而无穷”，认为自己掌控着世间真理，认为自己就是万物的法则，切实保障着社会的运行。比如马克思和阶级斗争、弗洛伊德和潜意识，或是马尔科姆·格拉德威尔和“引爆点”等。

而狐狸属于一种好斗的人，他们认为“千伎百俩而有尽”，解决问题有许多方法。他们对于琐碎、不确定、复杂或是有分歧的意见更加有耐心。如果说刺猬是猎手，总在不停地寻找大型猎物，那么狐狸更像是一个采集者。

泰特罗克发现，作预测时，狐狸型专家比刺猬型专家考虑得更周全，比如在苏联的问题上，他们的预测就更准确一些。对苏联问题进行预测时，他们没有把这个国家视为一个意识形态符号，既不认为苏联是“邪恶帝国”，也不认为苏联是马克思主义经济体系中一个相对成功（或是举世瞩目）的案例。他们只是客观地看待这个国家：一个濒临瓦解的国家，一个日渐式微的国家。如果说刺猬型专家作预测只能称得上是碰运气，那么狐狸型专家就可谓预测的高手了。

表 **2–2** 狐狸型专家和刺猬型专家的不同态度

狐狸型专家的想法	刺猬型专家的想法
“千伎百俩”：汇聚不同学科的思想，忽略最初的政治派别	“一技之长”：把大部分精力投入一两个重大问题上，以怀疑的眼光看待“局外人”的观点
适应力强：最初的方法失效后，试图找到新方法，或同时寻求多种方法	坚持力强：坚持“总揽一切”的方法，新数据只能用来改善原始模式
严于律己：有时会愿意（或是欣于）承认预测中的错误并接受谴责	固执己见：错误归咎到坏运气或特殊情况上——好模式没有赶上好时机
承认复杂性：承认宇宙的复杂性，认为许多基本问题不可解决或本身就是不可预测的	寻找秩序：一旦从噪声中找到信号，便期望世界遵循某种相对简单的支配关系
谨慎：用概率术语表达预测结论，并且证明自己的观点是正确的	自信：很少对自己的预测进行正面回复，并且不愿改变自己的预测
	意识形态：期待日常的问题正是宏伟理

经验主义：更多地依靠观察而非理论	论或斗争的体现
狐狸型专家是较好的预测家	刺猬型专家是较差的预测家

刺猬型专家更适合做电视节目嘉宾

一个冬日的午后，我在杜兰特酒店（加州大学伯克利分校附近的一家精品酒店）与泰特罗克共进午餐。他的表现足以证明自己是狐狸型专家：柔声细语，严谨治学，每次回答我的问题之前都会停顿二三十秒（唯恐给我的答复太仓促）。

泰特罗克问我："成为公众学者的动机是什么？一些学者非常低调、避开公众视线，但另外一些人则急切地想成为公众学者，高调、张扬、引人注目，这样才更有可能吸引众人的眼球。"

换句话说，高调、夸大的刺猬型预测更有可能使你成为电视中的公众人物。曾任美国前总统克林顿顾问的迪克·莫里斯现在是《福克斯新闻》的评论员，他是典型的刺猬型专家，他的策略就是抓住一切机会做出惹人注目的预测。2005 年，莫里斯称，小布什总统处理卡特里娜飓风的方式会使他重新赢得民众的支持。2008 年美国总统大选前夕，莫里斯预测奥巴马会赢得田纳西州和阿肯色州的竞选。2010 年，莫里斯预测共和党人会轻而易举地赢得美国众议院的 100 个席位。2011 年，莫里斯预测唐纳德·特朗普会竞选共和党候提名选人，并极有可能获胜。

然而，所有这些预言最终都落空了。卡特里娜飓风是小布什政府走向终结的开始，奥巴马在田纳西州和阿肯色州的竞选以惨败收场——实际上，奥巴马在这两个州的表现比 4 年前与小布什争夺美国总统宝座的约翰·克里的表现还要糟糕。2010 年 11 月共和党人确实有所收获，但他们只赢得了众议院的 63 个席位，而不是莫里斯所说的 100 个席位。莫里斯坚持认为特朗普会竞选共和党提名候选人，但在他做出预测后仅仅两周，特朗普就宣布退出了。

但是，莫里斯思维敏捷，说话风趣，非常善于推销自己——他在《福克斯新闻》中有固定的栏目，他所著图书的销量也高达几十万册。

狐狸型专家有时对刺猬型专家游走于电视节目、商业和政治活动的做法很难苟同。狐狸型专家认为，许多问题难以预测，所以我们应该对这些不确定性进行解释。他们的这种做派会让人误认为他们缺乏自信，他们的多元方法也被误解为缺乏确定性。杜鲁门总统曾经提出一个著名的论断，说自己需要一个"独臂经济学家"，他觉得那些狐狸型专家根本无法为自己的决策提供任何确定性的意见。

但是，狐狸型专家偶尔也会做出很好的预测。他们能较快地意识到

数据的嘈杂，而不去盲目跟踪错误的信号。与刺猬型专家相比，狐狸型专家更清楚自己还有很多无知之处。

如果你想让医生为你预测某种药物的疗效，或是想让投资顾问为你估算养老金的最大回报，那你最好选择狐狸型专家。他们可能会很谦逊地说自己的能力有限，但是成功的把握其实很大。

政治预测为什么常常失败？

在进行政治预测时，狐狸型专家的态度也许尤为重要。在进行政治展望时，刺猬型专家很容易落入陷阱，而狐狸型专家却能小心地绕过这些陷阱。

其中一个陷阱就是党派意识形态。尽管莫里斯一直以克林顿总统顾问的身份出现，却常被视为共和党人，并且为共和党候选人筹集资金，他保守的观点与《福克斯新闻》的风格也很吻合。但是，自由主义者却极有可能成为刺猬型专家。在对“麦克劳夫伦讨论小组”成员预测的精确性进行研究时，我发现，埃莉诺·克里夫特通常是小组成员中最自由的，她几乎从不发表对共和党特别有利的预测，总是跟小组里其他成员的总体意见相左。也许正是因为这一点，她才在 2008 年美国总统大选的预测中尝到了甜头，然而久而久之，她的预测也和保守派一样不准确了。

与泰特罗克的研究对象相似的那些专家学者，可能都会遇到同样的问题。实际上，对于拥有博士学位的刺猬型专家而言，一知半解是很危险的事。泰特罗克还有一些更重要的发现，其中之一便是，狐狸型专家的经验越丰富，预测就越准确，而刺猬型专家则恰恰相反：他们获得的额外信息越多，表现得就越糟糕。泰特罗克认为，刺猬型专家掌握的事实越多，他们以自己的方式改变且操纵事实的机会就越多，而这些方式往往会强化他们的个人偏见。如果你让一个忧郁症患者待在一个可以上网的黑暗房间里，就会出现类似的情况。你给他的时间越多，他要处理的信息越多，他最后得出的自我诊断也就越荒谬，不久之后，他就会把普通感冒误认为腺鼠疫了。

泰特罗克发现“左翼”和“右翼”两派刺猬型专家都会做出格外差劲儿的预测，但是所有政治派别中的狐狸型专家都不会出现这样的问题。狐狸型专家也许内心也会强烈地期待一种理想状况的出现，但他们在对现实情况和未来可能出现的情况进行分析时，通常可以将自己的理想与现实状况区分开来。

相反的，刺猬型专家总会将自己固有的偏好与分析的问题混为一谈。用泰特罗克的话说就是，他们创作了一个将“真相与个人价值观搅

和在一起的混合体”。他们对证据总是带有偏见，仅凭个人喜好取证，不尊重客观事实。

你可以用泰特罗克的测试为自己作个诊断，看看你是不是一个刺猬型的人：在获得更多信息的时候，你的预测准确率是否有所提高？理论上来说，拥有更多信息会让你的预测能力如虎添翼，不过你也可以忽略那些没用的信息。但是，刺猬型专家做不到这一点，他们经常会陷入信息荆棘之中，无法自拔。

《国家政治内幕期刊》（下文简称为《国家期刊》）曾经作过一项调查，对象是政治家、政治顾问、民意测验人和评论员，人数大约为 180 位。这项调查在民主党和共和党中分别展开，但是两个党派的受访人员回答的问题是一样的。抛开政治派别不谈，这组受访人员都比较接近刺猬型：那些政治活动家以身上的战争创伤为荣，并且发现自己一直都在与另一派别——鸡尾酒会派——进行无休止的斗争。

2010 年美国总统中期大选前期，《国家期刊》要求专门小组成员讨论民主党是否有可能继续掌控参众两院。大家几乎一致认为，民主党会掌控参议院，而共和党会掌控众议院（事后证明这两项猜测都是对的），两党的成员也基本同意由共和党获得众议院的大部分席位。民主党的专业人士呼吁本党派要拿到其中 47 个席位，而共和党则预测民主党将获得 53 个众议院席位，考虑到众议院中有 435 个席位，两党之间的预测差别可以忽略不计。

同时，《国家期刊》还要求专门小组成员对 11 场个人竞选的结果进行预测，包括参议院、众议院和州长竞选，但这一次的预测结果则颇有出入。在预测内华达州、伊利诺伊州和宾夕法尼亚州的参议员竞选、佛罗里达州的州长竞选和艾奥瓦州的关键众议院竞选中，小组成员的意见产生了分歧。整体上看，支持共和党的成员预计民主党只会赢得 11 场竞选中的一场，而支持民主党的成员则预计自己的党派会赢得 6 场竞选。（实际结果是两派预测的中间值——民主党最终赢得了 11 场竞选中的 3 场。）

很明显，党派偏见对预测产生了一定影响：民主党和共和党成员都会从自己的党派利益出发去考虑问题。然而，这并不足以解释为什么小组成员在回答不同类型的问题时会出现不寻常的分歧。当问及一般问题时，如共和党人可能会有何种表现，小组成员之间的答复几乎没有什么区别。但当问及具体问题时，他们的答案就会大相径庭，党派分歧就会显现出来。

刺猬型专家掌握太多信息可不是一件好事。共和党会赢得多少席位终究是一个抽象的问题：除非你仔细地研究了全部 435 场竞选，否则没

有多余的细节能帮你解决这个问题。相反，当问及某一次竞选时，比如内华达州参议员竞选，小组成员就需要处理各种各样的信息：这些信息不仅包括投票数，还包括自己读到的相关新闻报道、有可能是从朋友中打听到的小道消息，以及他们在电视上看到候选人时想到的。他们甚至与候选人相熟，或者认识候选人身边的工作人员。

刺猬型专家一旦掌握大量信息就会编一些故事，这些故事甚至比真实世界发生的事情更有条理、更加有序，有主角和反派、胜者和败者、高潮和结局，通常以大团圆结局收场。竞选票数低 10 个点的候选人将会取得最后的胜利，你若问刺猬型专家为什么会这样，他会说，因为“我”了解这个候选人，“我”了解她所在的那个州的选民，“我”还从她的新闻秘书那里听说票数咬得很紧。他还会反问你，你没看过她最新的商业广告吧？

当我们编故事的时候，就无法以批判的方式看待信息。选举通常都是扣人心弦的叙事性故事。无论你怎样看待巴拉克·奥巴马、莎拉·佩林、约翰·麦凯恩以及希拉里·克林顿这些人在 2008 年的政治表现，你都得承认他们的人生经历确实很有说服力。他们撰写竞选纪实著作，如《规则改变》，它读起来颇像一本畅销小说。参加 2012 年美国总统竞选的几位候选人虽然没有引起太多关注，但依然使民众过足了戏瘾，从赫尔曼·凯恩的悲剧到里克·佩里的闹剧，全部涵盖在内。

叙事性故事会使人忘乎所以。政治之所以特别容易受错误预测的影响，恰恰是因为其中的人为因素：每一次竞选我们都要投入很多情感。这并不是说，为了对某一政治事件做出准确预测，你对这件事就一定不能带有任何感情因素。但是这再一次表明，狐狸型预测者的冷漠态度更有助于他们做出准确的预测。

狐狸型预测方法

2008 年 2 月，我在新奥尔良的路易斯阿姆斯特朗国际机场候机，我的航班延误了。在等待登机的过程中，我萌生了建立 538 网站的想法。出于某种原因——也许是卡津马提尼酒在作祟——我觉得应该有人牵头建立一个网站，然后用这个网站来预测奥巴马和希拉里（当时两人还在为争夺民主党总统提名权激战）将如何完胜麦凯恩。这个想法突然间就明朗起来了。

我对政治选举的关注时间可能还要早一些，之所以关注，并不是因为热爱政治，而是对它感到失望。2006 年，我曾密切关注过美国国会下令禁止“互联网扑克”这一事件，“互联网扑克”可是我的主要经济来源之一。我发现，即使与体育报道相比，政治报道也依然存在盲区，前者

的报道和分析至少借着《点球成金》节目有所改善。

在大选之前，我发现自己越来越关注政治节目，主要收看美国广播电视台、美国有线电视新闻网和福克斯新闻的节目，可大部分内容都索然无味。尽管离美国总统大选还有数月之遥，但各档节目的评论都认为希拉里必然会获得总统提名权，却忽略了民意调查这种方式本身所固有的不确定性。这些评论似乎过于强调希拉里的性别优势和奥巴马的种族劣势，若某个候选人在一场新闻发布会上开了一个机智的玩笑或获得某个无名议员的认可，新闻界就会以此大做文章，断言这个人最终定会获选，可 99%的选民都不会在意这些事情。

政治新闻，尤其是那些真正影响美国总统大选的重要新闻，传播速度十分惊人。但是，新闻报道每天都有，且大部分都是用来填充版面的，而媒体却总是将这些报道打造成劲爆新闻。政治新闻报道不仅经常抓不住信号，还时常会强调噪声。如果一个州有大量选票显示共和党领先，而且大家都知道这件事，这就称不上是新闻。如果一个新的投票结果显示该州民主党处于优势地位，即使选票并非来自投票的主要群体，而且也不能准确预测出结果，但也一定会上头版头条。

换句话说，竞争设置的障碍当时已经相当小了。某些人只不过是对政治活动中真正拥有预测能力的事物进行了相当基础性的研究，他们看上去可能就像一个天才了。于是，我开始在 Daily Kos 网站上发表博文，公开了民意调查和募捐数目这类问题的细节，利用数据对这些问题进行分析。我研究过哪些民调公司在过去这些年的预测最准确，也研究过赢得某一个州（如艾奥瓦州）的选票究竟会对另外一个州的选票数产生多大的影响。尽管像 Daily Kos 这样的网站上的评论通常都是定性的（而且具有党派取向），而不是定量的（不用数据说话），但是我的那些文章还是很快就有了跟帖。2008 年 3 月，我将作过的分析转到我自己创建的 538 网站上，旨在对美国总统大选进行各种预测。

起初，538 网站的预测模型十分简单——基本上总是取自民意调查的平均值，但会根据之前的准确程度给这些调查加权——后来就逐渐变得比较复杂了，但预测模型始终遵循三大原则，而且这些原则都是典型的狐狸型原则。

原则一：用概率的方法思考问题

几乎我发布的所有预测，不管是政治方面还是其他方面，都是以概率的形式公布的。我不会随便给出一个准数或声称某事一定会发生，只会给出几个可能的结果。比如，2010 年 11 月 2 日，在预测共和党会赢得多少众议院席位时，我的估计正如图 2–1 所示。

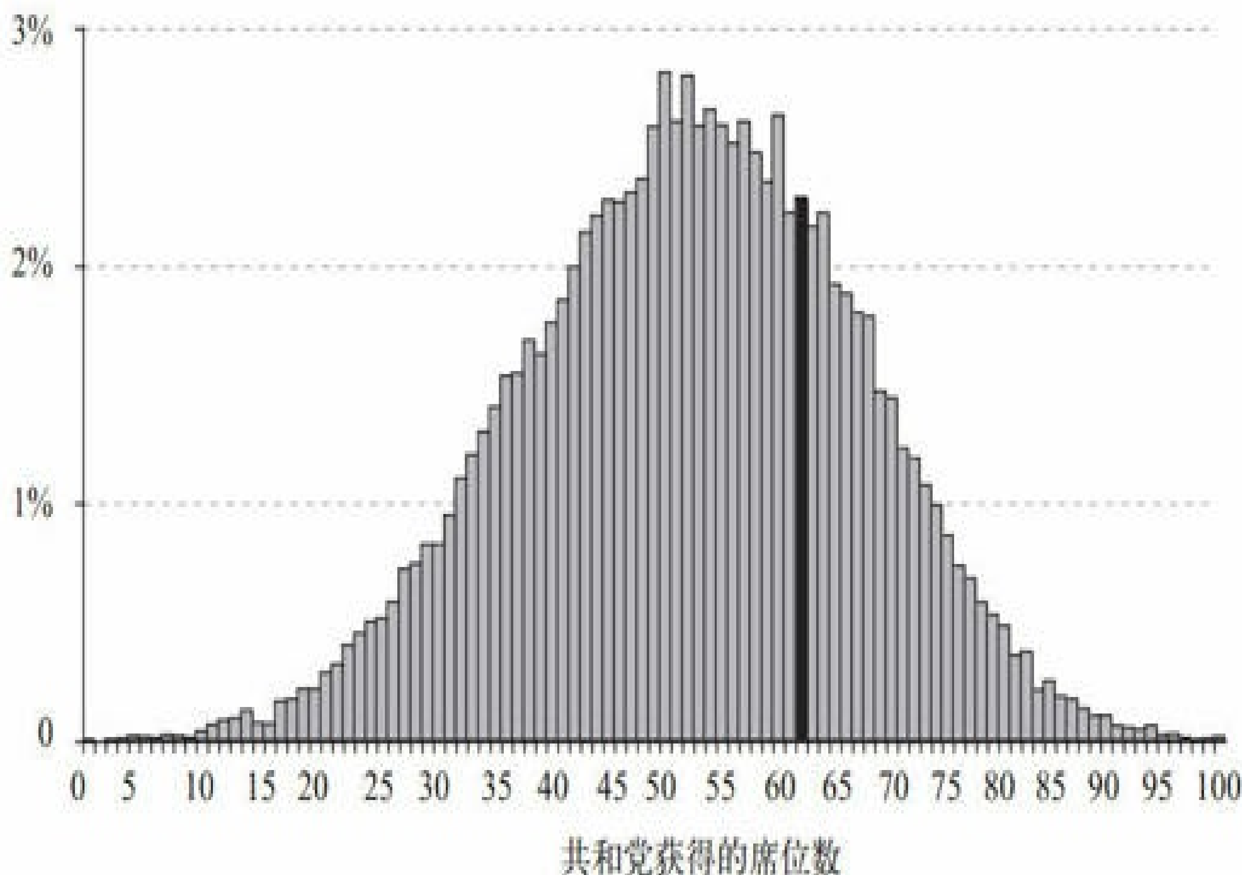


图 **2–1 2010** 年 **11** 月 **2** 日，**538** 网站预测模型对众议院席位的预结果测示意图

可能性最大的几个结果——足以覆盖所有可能出现情况的 1/2——显示，共和党会获得 45~65 个席位（实际上获得 63 个）。但是，也存在共和党会获得 70 或 80 个席位的可能性——这个数字与莫里斯当时预测的 100 个席位仍然相去甚远。而另一方面，民主党也有机会获得足够的席位以保住对众议院的控制权。

结果的广泛分布就是现实世界不确定性的真实体现。这个预测结果是在对 435 个席位中的每个席位分别进行预测之后得出的，其中多数席位的预测概率极为接近。尽管如此，仍有引人注目的 77 个席位的预测概率相互之间的差别大到个位数。如果当初民主党人在大部分有竞争力的地区能多赢得一些票数，哪怕是很少的一些，他们就能打败预测结果，轻松地保住对众议院的控制权。而如果共和党人采取了与他们的实际做法相反的做法，他们在众议院中的席位肯定也会低到让人大跌眼镜。政坛上，一颗小石子就会激起轩然大波，所以，预测时列出确切的数字肯定不是明智之举。

当我对某个竞选个案进行预测时，概率原则依然成立。比如，若民调显示某位候选人领先 5 个点，那他获胜的概率有多大？这就是 538 预

测网站试图解决的问题。

参选人能否获胜在很大程度上是由他参与的竞选活动的类型决定的。竞选级别越低，民调的不稳定性就越大：美国众议院选举的民调就不如参议院的精确，参议院选举的民调就不如美国总统大选的精确。初选的民调结果与最终大选的民调结果更是相差甚远。2008 年民主党初选期间平均票数丢了 8 个点，比预想的误差幅度大得多。2012 年共和党初选时民调的情况更糟糕。实际上，在许多主要的州，如艾奥瓦、南卡罗来纳、佛罗里达、密歇根、华盛顿、科罗拉多、俄亥俄、亚拉巴马和密西西比，民调中处于领先地位的候选人往往一周后就会落败。

但到了临近大选那几天，民意调查就会越来越准确。表 2–3 展示了 538 预测模型对参议院选举结果进行预测后做出的简化图，该模型运用 1998~2008 年的数据，根据美国候选人领先民调的平均得票情况推断其获胜的概率。比如，某位参议院候选人在选举前一天领先 5 个点，且以往他选举获胜过，那么他此次获胜的预测概率高达 95%，尽管新闻报道还会说这次竞选“难分伯仲”，但其实这次竞选他几乎已经胜券在握了。相反，如果是大选前一年领先 5 个点，那么他此次获胜的预测概率仅为 59%，这个概率跟掷硬币作决定差不多。

表 **2–3** 根据参议院候选人领先民调的平均得票情况预测其获胜的概率

领先情况				
距大选日期	领先一个点	领先 5 个点	领先 10 个点	领先 20 个点
一天	64%	95%	99.7%	99.999%
一周	60%	89%	98%	99.97%
一个月	57%	81%	95%	99.7%
三个月	55%	72%	87%	98%
六个月	53%	66%	79%	93%
一年	52%	59%	67%	81%

这样，538 网站的预测模型的价值就显现出来了。有了它，人们就可以非常轻松地对大选进行预测，观察一个候选人是否在全部或者大部分民调中都处于领先位置，判断他是否会成为最终获选的幸运儿（除了少数例外情况，这一假设还是正确的）。预测候选人获胜的准确概率其

实更难。我们的大脑处于高度探测模式，一直都在寻找信号，但同时我们也应该对得到的数据的干扰程度进行甄别。

对于这种思考方式我已经习以为常了，因为我的专业背景主要包括两大学科：体育运动和牌类游戏。浸淫在这两个领域中，你会明白什么叫见惯不怪。牌类游戏玩多了，你会轻而易举地抓到“皇家同花顺”，久而久之，当你的对手抓到一手“同花”牌时，你的牌可能已经满堂红了。体育运动，尤其是棒球，其中也有很多出现概率很低的事件最终实实在在地发生了。2011 年，波士顿红袜队当时有 99.7%的机会成功打入季后赛，可还是铩羽而归。对此我很无奈，只能说，一般的概率法则对红袜队和芝加哥小熊队不起作用。

然而，政客和政治观察员对这种不确定性却深感不满。2010 年，一个民主党的国会议员在美国总统大选前几周给我打来电话。他代表西海岸一个民主党占绝对优势的地区，但是考虑到共和党当年的良好表现，他还是很担心会丢掉自己的位置。他想知道的是，我们的预测会有多大的不确定性。我们为他提供了一个估值：民主党的获胜概率无限接近 100%。但近似 100%指的是 99%、99.99%还是 99.999 9%呢？如果他落选的概率只有十万分之一，他就应该把自己的竞选基金捐赠给民主党优势不够明显的地区的候选人。但现在一切都不确定，他连 1%的风险也不愿意冒。

同时，不同政治派别还会对你预测中的不确定性进行曲解。他们认为你是在两面下注，你在为可能出现的预测失误找借口。实际上不是这样的。如果你预测某位现任国会议员获选的可能性为 90%，你同时也是在预测他有 10%的可能性落选。一个出色的预测的标志是，每一个概率最终都会被事实证实是正确的。

泰特罗克所定义的刺猬型专家对概率的理解尤其差劲。当你说某个事件有 90%的概率会发生时，这句话本身拥有非常具体且客观的意思，但我们的大脑却会把这种说法转变为比较主观的推测。心理学家丹尼尔·卡尼曼和阿莫斯·特沃斯基给出的事例表明，这些主观推测通常并非总是与客观事实相吻合的。一架飞机安全着陆的概率是 90%、99%或者是 99.999 9%，我们看不出这 3 个数字有多大差别，但是在考虑应该订哪个航班的机票时，这 3 个概率对我们而言立即有了天壤之别。

多加练习，我们的预测就会变得更准确。泰特罗克所定义的那些刺猬型专家与其他人的不同之处在于，他们很固执，不愿意从自己的错误中吸取教训。想让他们承认自己对这个真实世界的预测存在不确定性，就相当于让他们承认自己对这个世界发展趋势的论断存在不足，这是所有理论家最不情愿做的事情。

原则二：今天的预测是你以后人生的第一个预测

另一个误解是：一个成功的预测不会改变。当然，如果你的预测结果每天都会出现巨大的波动，那可能就是一个不好的预示——不是设计的模型很差劲儿，就是你的预测对象根本不具备可预测性。2012 年，在各州公布共和党初选的预测结果之前，我抢在他们之前发布了自己的预测结果。我只是根据民调结果进行预测，这些概率经常会随着民调结果的变化而变化。

若结果的可预测性更强——比如美国总统大选到了竞选后期——预测结果通常会更加稳定。2008 年美国总统大选之后，我听到来自民主党最多的评论就是，他们向 538 网站求助以稳定党内人心。到了总统竞选后期，各个州每天会发布近三四十份民调结果，其中一些不可避免地会出现很大的偏差。那些喜欢把竞选气氛变紧张的候选人、战略家和电视评论员可能会把重点放在局外人的民意调查上，但是 538 网站的预测模型发现，这些民调不会对大选结果产生太大影响。

最根本的是，不管你上周、上个月还是去年说了什么，“今天做出最好的预测”才是应有的态度。做出新的预测并不意味着旧的预测不复存在了（最好对整个预测过程做个记录，并请人对你的预测效果进行评价）。但是，如果确定昨天的预测是错误的，那就没必要坚持了。经济学家约翰·梅纳德·凯恩斯有一句著名的论断：“当实际情况发生改变时，我的想法也会随之改变，你是怎么做的呢？”

有些人不认同这种进程–修正分析的方式，误认为那是软弱的表现。改变想法似乎是在欺骗自己——无异于伸出手指去测量风向。批评家通常含蓄或明确地认为，政治与物理、生物等学科类似，本身就应当遵循可知性和可预测性等基本法则。（我接触最多的批评家中有一位是普林斯顿大学的神经科学教授。）这样看来，新信息就不那么重要了，各种竞选就像冲向地球的彗星一样，应当沿着可预测的轨道行进。

然而，选举预测与物理和生物不同，倒是有点接近扑克牌游戏：我们可以观察对手的行为，从中找到一些蛛丝马迹，但是我们看不到他的牌。更新、更好的信息会不断出现，所以，只有时刻更新预测才能最大限度地利用有限的信息。因为害怕出丑而不对预测进行更改，这其实才是懦弱的表现。

原则三：寻求共识

每个刺猬型专家都幻想着自己能做出大胆、无畏又与众不同的预测——对一个话题的看法一定要跟众人的意见迥然不同才行。刺猬型专家遭受同事的排挤，就连家里的金毛猎犬也认为他们有些滑稽。但很快，

结果证明他们做出的预测完全正确、绝对正确、不容置疑地正确。两天后，他们上了《华尔街日报》的头版，同时附有一张照片，他们坐在杰伊·雷诺的沙发上，个个都像大胆而勇敢的先驱。

这样的预测可能偶尔是正确的。专家的共识也可能会出错——如果当初某个专家预测到苏联解体，那他一定会受到此生最大的褒奖。然而，幻想的情节终归难以成真。狐狸型专家——包括我本人在内——虽说不是墨守成规的人，但当我们的预测与竞争对手的大不相同时，我们总会感到忧虑。

许多证据表明，群体预测要比个人预测准确，其准确率随预测领域的不同通常会高出 15%~20%。这并不是说群体预测就一定是准确的（我们会在本书稍后的部分对这个主题进行深入探讨），但从多个角度考虑问题总会大有裨益。

泰特罗克曾经对我说："狐狸型专家通常会独立思考问题，而刺猬型专家则习惯与别人共同商讨问题。"他的意思是，狐狸型专家已经培养出一种仿效集体决策过程的能力。他们不再咨询整个专家组，而是不停地向自己发问。通常这就意味着他们需要将很多不同类型的信息集合起来——就像一组各持己见的人真正要做的那样——他们会客观地对待所有信息，而绝不会把某一条信息当成金科玉律。（以 538 网站的预测为例，这些预测通常会把民调数据和经济、各州人口统计等有关信息综合起来）。那些没有留心泰特罗克指导意见的预测者，总要为此付出代价。

2000 年美国总统大选前夕，经济学家道格拉斯·希布斯公布了一个预测模型，声称依据经济增长和战争伤亡率这两个变量，就可以非常准确地预测出总统大选的结果。希布斯提出了一些十分大胆的刺猬式主张，他认为依据总统的支持率(通常这是总统能否连任的可靠信号)，根本不会提高预测的准确性，对通货膨胀率和失业率有所了解同样也不会提高预测的准确性。候选人的身份无关紧要：一个政党既会推荐乔治·麦戈文那种有着强烈党派意识的议员，也会推选德怀特·D·艾森豪威尔那样的中间派战争英雄。希布斯断言，关键的因素是一个相对模糊的经济变量，即被命名为"实际可支配的人均收入"。

那么，希布斯的预测模型究竟得出了什么结果呢？它预测戈尔会获得压倒性胜利，并且说他最终会以 9%的优势取胜，但在佛罗里达州重新计票之后，获胜的却是小布什。戈尔确实赢得了美国普选，可这个预测模型表明，美国总统大选结果和预测结果会有天壤之别，且出现预测结果的概率只有 1/80。

还有其他几种预测模型也采用相似的方法，把美国总统大选这类复

杂的问题用“两个变量”的方程式解决（奇怪的是，这些模型中所用的两个变量五花八门）。实际上，在这些预测模型中，有些模型的跟踪记录远不及希布斯的预测模型准确。2000 年，其中一个预测模型预计戈尔会以 19 个点取胜，甚至还预测小布什只有十万分之一的概率获胜。

这样的预测模型在 1988 年美国总统大选之后开始盛行，因为在这次竞选中此类模型的基本变量一直偏向老布什一方——经济运行良好，老布什代表的共和党有一位颇受民众欢迎的前任总统里根——但直到本次大选后期，民意调查都是偏向迈克尔·杜卡基斯一方的。最终，老布什轻松获胜。

然而，尽管这类预测模型越来越多，但它们的跟踪记录却很差。1992 年以后的 5 次美国总统大选中，这个“基于基本变量”的典型预测模型——它忽略了民意调查，声称在没有相关信息的情况下仍可辨清选民的动向——结果漏掉了几位主要候选人之间高达 7%的点数差距。而采用狐狸式方法的预测模式，把经济数据、民调数据以及其他类型的信息结合在一起，得出了更为可靠的预测结果。

定性信息与定量信息同等重要

虽然这些“魔法子弹”式（结果发生逆转）的预测模式依据的是定量信息，比如已公布的经济统计数据，但还是失败了。实际上，本书所引用的那些最差劲的预测案例有些就是定量分析的案例。比如，那些评级机构就是运用不同模型对不同类型的抵押贷款的违约率进行预测，它们得到的都是精确到具体数据的估值。然而这些模型却漏洞百出，因为它们依据的是一种利己假设——不同抵押贷款之间的违约风险互不相干——这一假设在房地产市场和信用泡沫中完全行不通。当然，我自己在作预测时也非常喜欢运用定量的方法。刺猬型专家会接收各种类型的信息，并借助这些信息强化他们的偏见，而狐狸型专家则会对不同类型的信息进行总体权衡，将定性分析和定量分析结合起来，所以能经常做出正确的预测。

说到预测成功率，没有几个政治分析家能比得过“库克政治报道”这样一个紧密团结的团队。来自路易斯安那州的查理·库克长了一张友善的圆脸，他在 1984 年组建了这个团队，当时在贝尔特威以外的地方几乎没有人听说过他们。但是，政治狂热者多年来一直十分信任库克团队的预测，而且这个团队也几乎没让他们失望过。

库克和他的团队有一项特殊使命：对全美各级选举的结果进行预测，特别是对美国国会的选举结果进行预测。这就意味着，每隔一年，他们就要对外发布对美国众议院全部 435 场竞选结果的预测，还要发布

对美国参议院大约 35 场竞选结果的预测报告。

预测美国参议院选举或是州长竞选的结果相对简单。因为选民一般对这些候选人都非常熟悉，这些最为重要的竞选活动往往会引起广泛关注，还会有声誉良好的公司定期进行民意调查。在这种情况下，想要像我在 538 网站中那样改进收集民调结果的方法并非易事。

而美国众议院选举就完全是另外一番景象了。候选人通常身份比较低微，其中包括那些想在美国政坛一展身手的市议会议员或小企业家，而且在选举前，这些候选人一般不为选民所熟知。同时，美国国会选区几乎遍布全国各个角落，全美国人都被调动起来了。如果有人能够提供众议院选区的民调结果，这个结果必定是最起伏不定的，当然通常情况下几乎没有人能提供。

但这并不意味着像库克这样的分析家就无从知晓信息了，事实上，他们得到的信息量很大：除了民调结果，他们还知道某个选区的人口数据，也了解该选区此前几次选举的投票情况。他们还掌握了美国党派的总体发展趋势的数据，比如现任总统的支持率。此外，他们甚至能得到必须上报美国联邦选举委员会的筹款数据。

除了上述信息，其他类型的信息更偏向于定性信息，但同样很重要。比如，这位候选人是一个出色的公共演说家吗？她所在选区的政纲基调是什么？她的竞选口号是什么？政治活动其实就是一个小公司，她能管理好自己的职员吗？

当然，如果你是一个不仔细权衡定性信息的刺猬型专家，那么所有的信息就只会给你带来麻烦。库克的团队在预测方面很有经验，准确度也很高。

库克团队把竞选划分为从“共和党必胜派”到“民主党必胜派”7 个等级。1998~2010 年间，库克团队成功预测共和党候选人获胜的次数为 1 205 次（共 1 207 次竞选），达到了 99%以上的成功率，同样的，他们成功预测民主党候选人取胜的次数为 1 226 次（共 1 229 次竞选）。

库克提到的那些“民主党必胜派”获胜的选区几乎每年都是民主党获胜，“共和党必胜派”获胜的选区每年则都是共和党获选，这些预测都不难作。但是，在需要高技术水平的竞选预测中，库克团队同样表现不俗。比如，在被他们列为“倾向于”共和党候选人的那些选举中，共和党实际获胜的概率高达 95%；在被列为“倾向于”民主党候选人的选举中，民主党获胜的概率高达 92%。此外，即使库克团队并不赞同民调那样的定量指标，但他们在相关预测中仍然拥有良好的跟踪记录。

2010 年 9 月，大概是 2010 年 11 月选举活动开始前 5 周的光景，我在华盛顿拜访了库克团队，整个下午，我都和戴维·瓦瑟曼在一起，他

的年龄约为 30 岁，一头卷发，负责管理众议院的预测事务。

库克团队最具特色之处就体现在他们与候选人的面谈环节。美国总统大选期间，位于水门综合大厦 5 楼的库克工作室入口处总是人头攒动，候选人时不时地过来聊上个把小时，谈论的话题无外乎是筹款和战略会议。我拜访瓦瑟曼那天，他安排了 3 场面谈。他让我和一个叫丹·卡潘克的共和党候选人一起与他见面。卡潘克想要击败威斯康星州第三国会选区的现任民主党议员罗恩·坎德，这个选区包含这个州西南部的几个小社区。库克团队已经将这次竞选认定为“民主党可能获胜”，也就是说卡潘克获胜的概率很低，但是他们当时正在考虑将这一选区改为形势更有利的“民主党必胜”区。

做农场补给生意的州参议员卡潘克举止粗暴，像个高中体育老师。他还有一口浓重的威斯康星州口音，当他谈到自己拥有一支名为“拉克洛斯罗格”（La Crosse Loggers）的小型联盟棒球队时，我都不确定他说的是“logger”（伐木工），还是“lager”（一种淡啤酒的牌子），哪个词听上去都像是这个来自威斯康星州的俱乐部的别名。不过，他朴素的话语使人可以多少忽略其举止的粗暴，所以在通常民主党占优势的选区，他也始终能在州议会中任参议员。

然而，瓦瑟曼在面谈时运用了一些扑克牌游戏的规则。他一脸严肃且不失职业风范，但他会巧妙地给候选人施压，以找到更多信息。

他对我说：“我的基本技巧是，在面谈开始时让候选人谈谈自己的情况，从而营造一种舒适友好的气氛。然后，我试着提一些尖锐的问题，比如，说说你在哪些问题上与你所属党派的领导人意见不同。这么做的目的并不是让他们提供解决办法，只是想要了解他们的作风和处事方法。”

与卡潘克的面谈就是按照这一模式进行的。瓦瑟曼对选区的政治、地理状况无所不知，这使他看起来很像一个当地人，卡潘克也乐于谈论有关他所在选区的复杂情况，例如他在拉克洛斯需要获得多少票数才能弥补他在奥克莱尔丢掉的选票。他啰啰唆唆地说了几个问题，主要内容是他用说客们游说得来的捐款为俱乐部买了一个全新的棒球场。

这都是一些小事，远不及卡潘克因为婚外情和偷税漏税被起诉来得重要。这两件事足以劝阻瓦瑟曼不要改变评价。事实上，尽管那年 11 月共和党在中西部相似选区的大部分竞选中都取得了胜利，但卡潘克本人还是以 9 500 张选票的差距落选了。

事实上，这种情况很常见。面谈之后，瓦瑟曼通常会保持（与面谈前）同样的评价。尽管他努力地从候选人那里收集新信息，可还是不能推翻他之前对竞选结果的看法。

瓦瑟曼的方法之所以奏效，是因为他能在不受对面这位候选人影响的情况下，对信息做出评价。能力不足的分析师不是表现得像是被迷惑、被欺骗、被耍弄，就是在竞选活动中迷失自我。或者，他们只顾着对自己与候选人谈话时所使用的技巧感到沾沾自喜，却忽略了其他所有与竞选相关的信息。

而瓦瑟曼会考虑更大的政治背景中的每一件事。极其优秀的民主党候选人，即使面谈时表现超级棒，可能也没有机会在共和党通常获胜的选区取胜。

那为什么还要费时费力地与候选人进行面谈呢？这主要是因为瓦瑟曼在寻找危险信号，就像那次与民主党议员艾瑞克·马萨（在传出性丑闻之后突然从国会辞职）面谈一样。心理学家保罗·米尔将这些情况称为“断腿”情况，即如果一件事情太过招人耳目，那么不对它进行解释说明就是不明智的。

这种现象，每年瓦瑟曼都会遇到几例，这使得他能够正确预测更多的竞选。他能够从面谈中得到足量但不过量的信息。信息是定性的还是定量的，这并不重要，重要的是你怎样使用这些信息。

做出客观的预测并非易事

在本书中，我使用“客观”和“主观”这两个词时，都非常谨慎。“客观”有时被认为是“定量”的同义词，其实不然。它的实际意义是超越我们的个人喜好和偏见，去寻找问题的真相。

在这个世界上，纯粹的客观性是求之而不得的。进行预测时，我们可以选择很多不同的方法。有些方法可能单纯依赖定量因素，比如民意调查，但是瓦瑟曼的方法也将定性因素考虑在内。然而，所有这些决定和假设都要由预测人做出。人为的判断终究会存在潜在的偏见。只有认识到假设对预测的影响，并从自身找问题，才能做到更加客观。在政治预测上，我们穿梭于意识取向和行为倾向之间，在嘈杂的数据中理顺客观事实，这是相当困难的。

所以，你需要从电视评论员身上学习一些不同的习惯，学会怎样表达预测中的不确定性，并使其量化。随着事实和环境的改变，你的预测也要跟着更新，你要试着发现，从不同视角看世界也会得到智慧。只有这样，评估各式各样信息的能力才会越来越强，而不至于滥用这些信息。

简而言之，你需要学习狐狸型专家的思维方式。狐狸型专家认识到了人类在预测世界进程时所作判断的局限性。认识到这些局限，才能做出更准确的预测。

第三章
棒球比赛预测：球探和数据怪才，谁更胜一筹？

刚从纽约回来的红袜队士气非常低落，连输 3 场对阵劲敌洋基队的周末赛，使得红袜队与 2009 年美国棒球联盟东区冠军彻底无缘了。常规赛只剩 7 场了，红袜队似乎就要成为美国棒球大联盟季后赛的外卡队，而这样的季后赛绝不是主办方所期待的。有统计研究表明，球队常规赛的表现并不影响其季后赛的发挥，但 2009 年红袜队似乎已经放弃希望了。

在芬威球场，我准备采访红袜队的明星二垒手达斯汀·佩德罗亚。佩德罗亚是 2006 年以来我最喜爱的球员之一，2006 年我为年刊《棒球规程》创建的预测系统 PECOTA 也预测他将成为最佳棒球运动员之一。但这一预测结果却与球探的观点截然相反，他们认为佩德罗亚天资不足，身高欠缺，挥棒姿势怪异，只能当个替补球员。然而，在 2006 年最具发展前景的棒球运动员名单中，PECOTA 预测系统却将佩德罗亚排在第 4 位，而向来注重球探观点的出版物《美国棒球》则将他排在第 77 位。其实，关于佩德罗亚的报道中，这种褒少贬多的说法屡见不鲜。娱乐与体育节目电视网的基斯·劳早在佩德罗亚的新秀赛时就说过他没什么发展前途了。

> “达斯汀·佩德罗亚毫无实力可言，击球欠缺速度和力量，在顶级联赛中，也不能连续击投球。如果他的击球率能达到 0.260，还是有些价值的。但他总是连滚带爬地跑到三垒或游击手那里，如果能改掉这个缺点，成为一名内场替补球员还是有希望的。”

2007 年 5 月 12 日，基斯·劳发表了上述评论。那天，佩德罗亚的击球率只有 0.247，只完成了一次本垒打。说真的，当时我对他也开始失去信心了，我看了他大部分“安打”的比赛，佩德罗亚表现出一脸败相。

但是，佩德罗亚仿佛想证明那些怀疑他的人是错的，他开始发狠地击球，期盼迎来逆转。接下来的 15 场比赛中，他的击球率达到惊人的 0.472，使他的平均击球率从 4 月低谷时的 0.158 一路升至 0.336。

2007 年 7 月，也就是基斯·劳做出那番评论的两个月之后，佩德罗亚入选美国联盟全明星队。10 月，他帮助红袜队赢得有史以来第二个世

界职业棒球大赛冠军（上一次是在 1918 年）。11 月，他被评为“年度最佳新秀”。在随后的一个赛季，24 岁的佩德罗亚获选“最有价值球员”，一举成为美国棒球联盟中表现最全面的球员。佩德罗亚告别内场替补，成为今日巨星，而那些球探当日却严重地低估了这位棒球天才。

之所以到芬威球场来，是因为我想了解究竟是什么让佩德罗亚实现了逆转。我准备了一堆问题，也获得了红袜队的采访许可，可以进行场地采访。但我知道这并非易事，因为联盟球场是球员心中的圣地，在此地采访并不适合，且红袜队又刚输了周末赛，队内气氛乖戾紧张。

我在那里看着佩德罗亚进行场内练习，他接住大块头三垒手凯文·尤基里斯的投球，转投给新的一垒手凯西·柯奇曼。他显得与众不同，动作认真到位，而尤基里斯则略显笨拙，柯奇曼似乎还在走神。佩德罗亚与队友的主要不同之处在于态度，队友走神时，佩德罗亚会愤怒地将球摔出场地，恼怒地看着他们。

15 分钟训练过后，红袜队出发到当晚的对手多伦多蓝鸟队的主场。我站在离红袜队的休息棚一两米远的一垒位置旁边，佩德罗亚从我身边走过，正如那些球探所说，他的确不算高，按照官方公布的数据，他身高为 1.75 米（5 英尺 9 英寸），相当于把我的身高四舍五入之后的高度，但我好歹还比他高出两英寸。球探们说他不像一个运动员也没错，刚刚 25 岁就秃顶，头顶的头发和下巴上的胡须一样多，腰间赘肉堆积，还有小肚腩。如果你在街上遇见他，很可能会把他当作音像店的店员。

佩德罗亚转身走进休息棚，独自坐在那儿。这似乎是一个采访的好时机，我鼓起勇气走上前去。

“嘿，佩德罗亚，有空吗？”

他狐疑地盯着我看了几秒钟，然后用一种极尽优越的口吻，一字一句强调道：“不，我没空。我正在努力备战棒–球–大–联–盟–比–赛！”

于是，尴尬的我只能在球场边转悠，努力在前往记者席看比赛之前重拾自己的尊严。

第二天，采访许可到期之后，我回到纽约，委派从前的同事（《棒球规程》的资深记者）戴维·劳里拉前去打探，希望他能获得一些有价值的信息。但是，这次佩德罗亚的话仍不多，回答也是尽可能地平淡无奇。他告诉劳里拉：“你知道吗，我根本不在乎数字和统计数据，我只在乎输赢，其他的东西对我来说都无所谓。”

佩德罗亚曾经偏离政党路线，因此惹了一身麻烦，之后他就学会了用这种说话方式来应对各种提问。比如那次，他称自己的家乡加利福尼亚伍德兰是一个垃圾场，还告诉《波士顿杂志》，“你可以说这句话出

自我，事实上也是如此，我从不撒谎。”

佩德罗亚确实从不说谎。我想，也许正是因为持有这种态度，他才能够承受得起“球探”的质疑，最终成为大联盟中的一员。

构建棒球比赛的预测系统

从记事起，我就是一个棒球迷，也是一个棒球数据迷。6 岁时，我家乡底特律的老虎队赢得了 1984 年的世界棒球联赛，作为一个数学小神童，当时我对比赛的各种数字特别着迷。7 岁时我第一次买棒球卡片，10 岁时第一次读《埃利亚斯棒球分析师》，12 岁时有了自己的统计数据。（令人难以置信的是，那个数据竟然预测到默默无闻的红袜队场内球员提姆·尼尔林将成为最佳球员之一。）

直到 2002 年，我对棒球和棒球数据的兴趣才达到最高点。迈克尔·刘易斯当时正忙着写《点球成金》，书中记录了奥克兰竞技队及其总经理、统计奇才比利·比恩崛起的故事，该书很快畅销全美。25 年前，比尔·詹姆斯出版了《比尔·詹姆斯棒球摘要》（下文简称为《摘要》），开创了棒球数据统计分析时代，随后他被聘为红袜队的顾问。所有这些都让我发现，沉迷棒球数据似乎不再只是一个小小的爱好，它为我寻觅新工作提供了一个新思路。

大学毕业之后两年，我一直在芝加哥的毕马威会计师事务所担任转让定价顾问。工作不错，薪水不薄，老板和同事做事专业、待人友善，生活还算安逸。

平时，我要么指导某公司的马来西亚移动电话分部如何定价，以便降低税金，要么赶早上 6 点的飞机到圣路易斯为一家煤矿公司的合同估价，但这份工作并不是我想要的充满刺激的工作。当时我 24 岁，不愿意安于现状，对我而言，这份工作平淡无奇、按部就班、了无生趣，让我觉得生活乏味至极。但这份工作有一个好处，那就是我有很多空闲的时间可用。于是，闲来无事，我就开始创建各式各样的电子表格以记录庞大的棒球统计数据，这后来便成了 PECOTA 预测系统的雏形。

读大学时，我就开始看年刊《棒球规程》。这本刊物是加里·赫卡贝于 1996 年创办的。一头红发的赫卡贝热情奔放，嘴不饶人。他先是从早期互联网新闻组中招募了一批写手，后来成了体育运动数据分析行业的领军人物。加里·赫卡贝很早就嗅出了商机：1988 年比尔·詹姆斯停止了《摘要》的出版，而那些试图取代《摘要》的出版物不是品质不好，就是在 1994~1995 年棒球罢赛期间停刊了。于是，他在 1996 年创办了《棒球规程》刊物。第一本刊物问世时，是由激光打印机一次一本印出来的，还意外漏掉了圣路易斯红雀队的内容，最终《棒球规程》只售出

了 75 本。但之后这本刊物销量逐年倍增，成了热门刊物。

《棒球规程》是统计怪才的快乐源泉。数据海洋里不仅有大联盟球员的数据，还有小联盟中表现出顶级水平的球员的数据。刊物虽然有时只适合圈内人士传看，但文风犀利，常引用《辛普森一家》或是 20 世纪 80 年代情色电影里的隐晦笑话，还常对不受欢迎的联盟经理冷嘲热讽，读来颇为有趣。

当然，刊物中最重要的内容还是对每位球员下个赛季表现的预测，预测依据的是赫卡贝创建的“弗拉迪米尔”预测系统。这一系统似乎是继詹姆斯开创的棒球预测革命之后的又一个里程碑。

一个好的棒球预测系统必须可以完成 3 项基本任务：

1. 考虑影响球员表现的外在因素。
2. 区别看待技术和运气因素。
3. 熟知老化曲线，了解球员的表现是如何随着年龄增长而发生变化的。

任务 1 相对简单。在美国主要的运动项目中，棒球很独特，比赛场地没有标准尺寸。道奇体育场如血管般错综复杂，外围是壕沟般的停车场，相比之下，芬威球场则如缩略版的新英格兰网格街道，整齐、舒适，球员在此更容易获得较高的击球率。通过观察球员在主客场的表现，我们就可以用“球场因素”来解释球员所面临的困难的程度。（比如说，在弗莱德·林恩——20 世纪 70 年代红袜队的“最有价值球员”——的职业生涯中，他在芬威球场时击球率达 0.374，但在其他体育场击球率只有 0.264。）同样的，通过观察从国家联盟转会到美国联盟的那些球员的表现，就能分辨出哪个联盟更好，也能看出一个球员的竞争实力。

全世界最丰富的统计数据库

任务 2——区别看待技术和运气因素——就不是那么轻松了。在短期内，运气常会主导棒球比赛：即使最好的球队也会输掉 1/3 的比赛，即使最好的击球手安全上垒的概率也只有 2/5。有时候，运气因素甚至会掩盖一个队员的真实技术水平长达一年之久。某一个赛季中，一个击球手的真实击球率为 0.275，受运气因素影响，这个球手的击球率有10%的概率出现浮动，时而升至 0.300，时而降到 0.250。

设计完善的预测系统能够辨识出那些容易受运气影响的数据，比如，平均击球数就比本垒打数善变。投手的表现也是出了名的不稳定，所以数据整理对预测投手的表现至关重要。若想预测一位投手的输赢，

参考他取得的三振出局数以及保送数目，要比翻阅他前一赛季的输赢纪录更有价值，因为从每年的情况来看，前一类数据要更稳定些。

像所有预测一样，对棒球比赛进行预测是为找出决定输赢的根本原因：将击球手三振出局就能阻止对手安全上垒，从而阻止对手得分，进而阻止他们赢得比赛，所以三振出局数更为关键。然而，预测这回事，越往下探究，就会遇到越多的噪声干扰：投手的输赢纪录受自己发挥的影响，也受制于对手得分的多少，而后者是他无法左右的。西雅图水手队的明星投手菲利克斯·埃尔南德斯在 2009 年和 2010 年的投球表现都很不错，但输赢纪录却有很大差距，分别是 19∶5 和 13∶12，因为 2010 年水手队的击球手水平极差，给了对手很多机会。

此类事例的出现绝非偶然，只要用心整理数据，就会发现端倪。棒球比赛为我们提供的数据也许是全世界最丰富的：过去的 140 年里，在大联盟球场上发生的每一件事都被忠实、准确地记录下来，而且每年还有数百位棒球手在很多较大的联盟中打球。另外，尽管棒球是团队运动，却是以高度有序的方式进行的：投球手轮流投球，击球手按序击球，这样每人都有各自的统计数据。复杂的、非线性的统计问题很少出现，因果关系很容易理顺。

这使得棒球预测者的工作变得十分轻松。（在这项运动中，）一个假设通常可以根据经验进行检验，利用精确度比较高的统计数据来证实或者证伪。而对于经济预测或政治预测来说，可用的数据就少很多——美国总统大选每 4 年才举行一次，不是每年都能产生几百个数据点——缺少数据，预测往往更易误入歧途。

老化曲线与相似分数

前述观点都是建立在一项假设的基础上的，即球员的技术水平一直保持稳定，年复一年，不会改变。其实不然。只要能将信号与噪声区分开来，我们就能获得所需的任何信息。事实上，每位棒球队员的技术水平都处于一种不断变化的状态，预测的挑战也正在于此。

研究过数千名球员的统计数据之后，詹姆斯发现，在二十八九岁之前，大多数球员的技术水平都会持续提高，但之后通常会开始走下坡路，三十五六岁时尤其严重。詹姆斯由此得出他最重大的发现：老化曲线。

奥运会体操运动员的运动巅峰期是十几岁，诗人的黄金时代是二十几岁，国际象棋选手的最佳比赛年龄是三十几岁，应用经济学家在四十几岁时成就辉煌，财富 500 强首席执行官的平均年龄为 55 岁。詹姆斯发现，一个棒球运动员状态最好的年龄是 27 岁。1985~2009 年间，获

得“最有价值球员”称号的运动员共有 50 位，其中 30 人获此殊荣的年龄是在 25~29 岁，更有 10 人的年龄刚好是 27 岁。这个年龄是体能和脑力完美结合的最佳阶段，可以让棒球运动员发挥出最高水平。

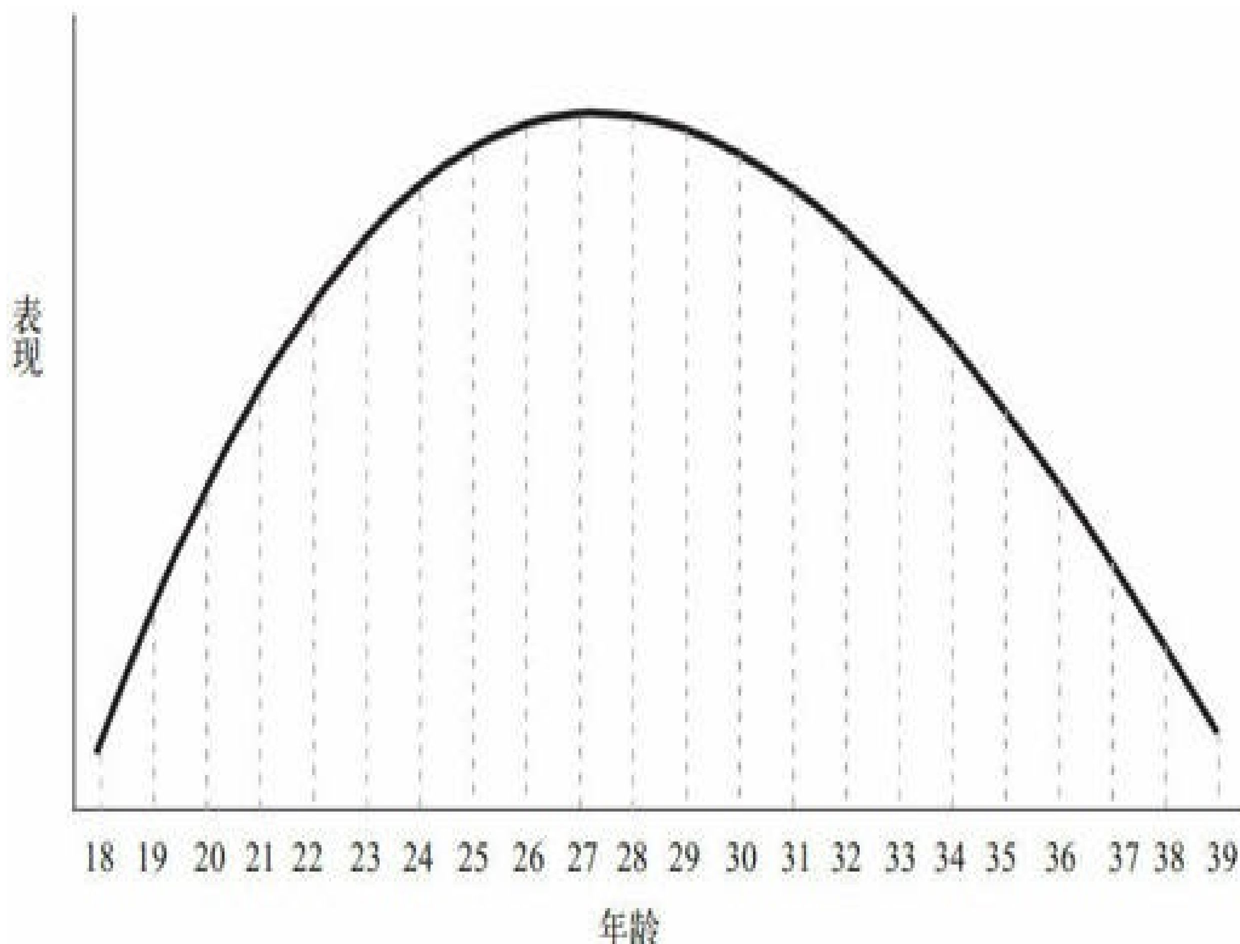


图 **3–1** 击球手的老化曲线

读过詹姆斯的著作的球队都能从“老化曲线”中受益良多。棒球合同规定，球员必须打满 6 个大联盟赛季方可成为自由球员，到那时已处于其职业生涯的后期了。（在此之前，所在俱乐部享有对他们的绝对支配权，而且他们的要价要低于市场价格。）通常，一个新晋球员会在二十三四岁时加入大联盟，直到 30 岁时他才可能成为自由球员——那时他的巅峰表现期已过，运动生涯开始黯然失色。球队为这些自由球员支付大量的保险费用，指望他们可以在 30 岁时复制二十几岁时的辉煌，而事实上球员的技术水平已不复当年，但大棒球联盟合约规定必须支付这笔费用，球队也没有办法。

詹姆斯的“老化曲线”并非没有缺陷，表现为其走势过于平缓。诚然，一般球员会在 27 岁达到巅峰状态。但只要是付过会费的人盯着棒球卡背面研究一会儿都会发现，不同球员的老化速度是不一样的。比

如，20 世纪 80 年代亚特兰大勇士队的三垒手鲍勃·霍纳获得年度最佳新秀奖时只有 20 岁，24 岁时就入选全明星队了。如果根据当时通用的假设预测，鲍勃进入名人堂就是板上钉钉的事。但 30 岁时他却因伤停赛了，之后又草率地加入日本大联盟的“养乐多燕子棒球队”，彻底结束了他的顶级职业棒球生涯。再如，来自西雅图水手队的伟大球员埃德加·马丁内斯直到 27 岁才在大联盟中有了一份稳定的工作，大器晚成的他在 30 多岁时才迎来了职业生涯的巅峰时期，40 岁时击球数还排在联盟的领先位置。

尽管鲍勃和马丁内斯的例子有些不合常规，但确实很少有球员如那条老化曲线般平缓发展，高峰低谷间断交错的形式反而更为常见。也就是说，真实的老化曲线充满噪声，而且噪声非常大（见图 3–2），不会那样平缓。平均值就像“每个家庭拥有 1.7 个孩子”这样的说法，不过是一个统计学的抽象概念。按加里·赫卡贝的推论来看，也许詹姆斯的老化曲线没能传递出噪声中的某些信号。也许像游击手这样处于对体能要求比较高的位置的球员，其技术水平的衰退速度要比打右外场的球员快。也许那些颇具运动天赋的全能型球员的职业生涯会比那些只有一两个强项的球员要长。

赫卡贝的系统则假定存在 26 条明显的老化曲线，每一条都适用于不同类型的球员。如果他的假设是正确的，你就能评判出哪条曲线适合哪个球员，并以此预测出该球员职业生涯的走势。如果一个球员符合鲍勃·霍纳的老化曲线，那就预示着他会少年得志，但衰退期也会早早到来；如果一个球员符合马丁内斯的老化曲线，那就预示着他将大器晚成。

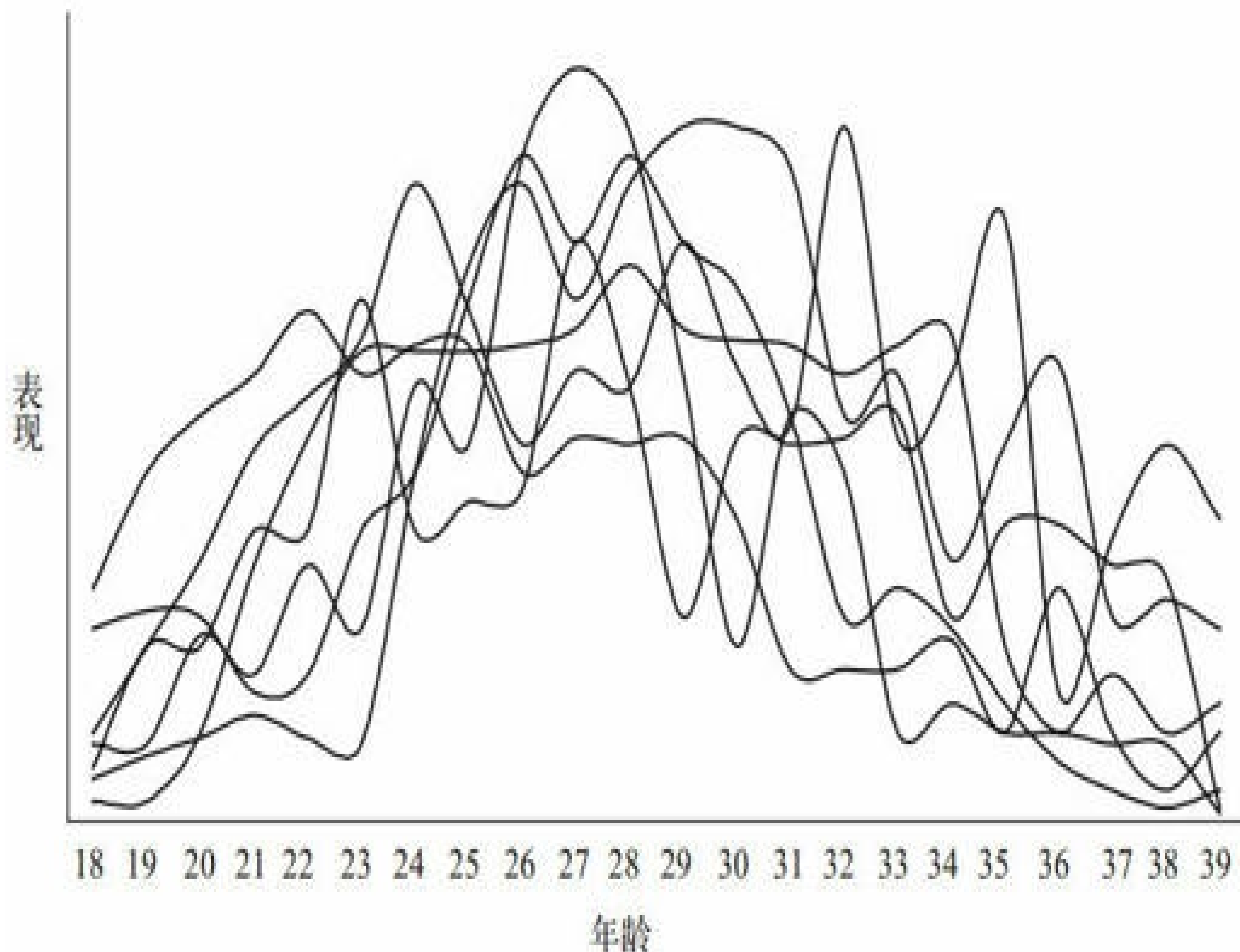


图 **3–2** 不同击球手的带有噪声的老化曲线

在赫卡贝的“弗拉迪米尔”预测系统中，每个球员对应着不同的曲线；而在詹姆斯创建的平缓的预测系统中，所有球员都对应着同一条曲线。尽管如此，从赫卡贝发布的一些预测结果看，“弗拉迪米尔”系统的预测准确度比詹姆斯的老化曲线高不了多少。预测准确率不高，在一定程度上是因为赫卡贝设定的 26 条曲线没有科学依据，只是随意选定了这个数目。这样一来，为球员匹配曲线就成了一个科学和艺术并重的技术活。

然而，要将棒球队员的技术水平发挥到较高的程度，一个人一定要对自己的身体技巧和精神技能进行多种多样的排列组合才行：肌肉记忆、体能耐力、手眼协调性、击球速度、投球识别力以及失利时强大的意志力，这些因素都要科学合理地运用。如此看来，“弗拉迪米尔”预测系统关于不同老化曲线的概念貌似更符合人类行为固有的复杂性。在创立 PECOTA 预测系统时，我尽量从詹姆斯和赫卡贝的系统中借鉴了一些元素来完善自己的预测系统。

在 1986 年的《摘要》中，詹姆斯介绍了“相似分数”的概念。顾名思

义，设计者设计了这些分数，用来对任意两名大联盟球员的职业数据的相似性进行评定。这个概念比较简单。评定人员首先将两名球员分为一组，每组赋值 1 000 分，每发现两人之间的一次不同则扣若干分。结果，高相似度的组可得到 950 分甚至 975 分，但低相似度组的组员间的差异会迅速增大。

略懂棒球历史的人都会发现相似分数的方法很得人心，评定人员并没有一门心思地只顾研究球员的统计数据，而是提供了一定的历史背景。例如，佩德罗亚在 25 岁时的数据与历史上的很多球员相似，比如 20 世纪 70 年代统帅明尼苏达双城队的队长、一个伟大的巴拿马人罗德·卡鲁，以及大萧条时期老虎队的明星查利·格林格。

詹姆斯主要是想用相似分数进行反向检测，比如分析一个球员是否有资格入选名人堂。如果想论证你最喜爱的球员是否够资格入选库佰斯顿球员名册，就可以借用相似分数。若通过观察你会发现已经入选的球员中有 9/10 的人的数据和他的相似，你就得到了一个非常有力的论据。

相似分数也可以用于预测吗？假设我们能够确定有 100 名球员某一个年龄段的数据与佩德罗亚相同年龄段的数据极为相似，那么根据这 100 名球员的职业生涯的整体表现难道不能预示佩德罗亚的职业生涯走势吗？

为了验证这个想法，2002 年我还在毕马威工作时就开始利用闲暇时间慢慢地推进自己的计划——创立 PECOTA 预测系统的雏形。这一系统是由五花八门、内容丰富的 Excel 电子数据表格构成，而表格正好也是我在毕马威工作时使用的主要工具（所以每次那些老板经过我的座位时，都以为我正在努力为客户建立细致的表格呢）。

就这样，上班时忙里偷闲每天抽出一两个小时，晚上到家再奋斗几个小时，我最终建成了一个数据库，其中包含 10 000 多个赛季（包含了自第二次世界大战开始后的所有大联盟赛季）以及所有球员的对比演算数据。这一方法充分利用了棒球比赛无与伦比的丰富数据，从某种程度上讲，它比詹姆斯的方法更细致。在对比一组球员时，此法采用最近邻体分析法，另外，考虑的因素外延更广，甚至包括球探经常关心的球员的身高、体重等因素。

PECOTA 预测系统为不同类型的球员提供了不同的老化曲线，这一点与赫卡贝的系统相似，但 PECOTA 并不局限于 26 种曲线，相反的，它从棒球庞大的数据库中识别出一组数据相似的球员，综合分析后自然生成了这组老化曲线。如果与佩德罗亚数据相似的球员最后都成了大联盟中的优秀队员，这就预示着他成功的概率也比较高。

但通常，那些拥有相似数据的球员各自又是一个复杂的混合体，其老化曲线也会在某个时间点开始相互偏离。之前提到，依据詹姆斯的相似分数方法，佩德罗亚的数据先是与格林格和卡鲁相似，后两位球员在其漫长的职业生涯中战绩显赫，入选了名人堂。但此后佩德罗亚的数据又与蒙特利尔博览会队实力平平的二垒手乔斯·维德罗极为相似了。

对小联盟中的球员而言，这种差异更为明显。2009 年，PECOTA 预测系统为亚特兰大勇士队 19 岁的“明日之星”杰森·海沃德识别出的最相似球员，竟包括从名人堂球星到谋杀犯球员。奇珀·琼斯与海沃德的相似度排名第一，这位名人堂球星、前亚特兰大最伟大的球员之一为俱乐部效力 17 个赛季，职业生涯里的平局击球率为 0.304，外加 450 次本垒打。而达奈尔·斯坦森是第二个与海沃德相似的球员，他的结局并不好，本来颇具潜力，却在 2003 年结束了在发展联盟亚利桑那队的球员生涯后，因屡次犯罪而被追捕，后来开着一辆运动型多用途汽车逃亡。

所有与海沃德相似的球员都是高大的运动全才，在小联盟比赛中展现出高超的技术水平，在选秀中脱颖而出，但是他们的命运却截然不同。而 PECOTA 预测系统的创新点就在于此，它能够告知球员他们可能遇到的各种命运结局，这些结局参考了数据相似的前辈球员，尤其是那些结局最好与最坏的球员，或结局最有可能再次上演的球员。由此可见，若预测对象是人类行为，那结局更是不胜枚举了。

看到现在，海沃德的命运真可谓跌宕起伏啊。而真实情况是，2009 年海沃德荣获年度小联盟“最佳球员”称号，2010 年他在自己前 30 场大联盟比赛中贡献了 8 个本垒打，完成了自己在勇士队的首秀，还入选了全明星队，这一切都超过了人们的预期。但是，2011 年海沃德在大联盟的第二个赛季却不是那么顺利，击球率只有 0.227。面对海沃德在 2011 年赛季的糟糕表现，完善的预测系统仍会找出一个保持乐观的理由：他还是有潜力的，除了平均击球率之外，主要数据仍然保持了先前的水平，而平均击球率本身就比其他数据更容易受到运气因素的影响。

但是，你想要了解的关于一个球员的一切，这些数据都能够告诉你吗？10 年前，这个问题可是棒球界最热的话题。

球探与数据怪才的矛盾冲突

只要对《点球成金》稍有了解你就会知道，书中讲述了“统计专员”与“球探”这两个敌对阵营之间的冲突与斗争，他们注重不同的范例，采取不同的方法评估球员的表现。（“统计专员”当然是用数据说话，而球探也有自己的“工具”。）

2003 年，《点球成金》一书出版后，迈克尔·刘易斯的读者们一定不

会忽略书中对两个阵营敌对状态的描绘（而这本书本身很可能也加重了两者间的敌意）。同年，我参加了在新奥尔良万豪酒店举行的棒球冬季会议，当时的情形一下把我带回了高中时代。一方是老牌球探，他们就像草地上的老水牛一样悠闲自在，偶尔会坐在酒店吧台旁一边喝着威士忌，一边聊着老掉牙的战时故事，更多的时候是把自己关在宾馆的房间里谈生意。这些人从前都是专业从事棒球运动的，而今多半都已届不惑之年甚至是知天命之年了。他们中很多人以前都是运动员，在球场上为棒球运动奋力打拼，后来逐渐进入领导层负责幕后工作，一路平步青云。另一方则是一群数据怪才：二三十岁的模样，背着电脑包，拿着彩色打印的建议书，不停地在大厅的人群中穿梭，希望可以说服某个投资者为自己提供一个工作岗位。平时这两个阵营的交流并不多，他们都觉得对方自大无知且闭目塞听。

两方矛盾的根源也许就在于，球探觉得数据怪才威胁到了他们的饭碗。他们认为球探预算正在缩减，不管这是客观事实还是主观感受，他们通常都会以此为依据，说自己的饭碗难保。在一次围绕《点球成金》展开的激烈讨论的圆桌会议上，洛杉矶安那罕天使棒球队的球探总监埃迪·贝克告诉《美国棒球》杂志的记者："此刻真是剑拔弩张啊。我们的一些老伙计本不该丢掉工作，现在却失业了。也许有财政或是其他因素的影响，但我们都认为其实是一些懂电脑的人抢走了他们的饭碗，对此我感到十分愤怒。"

我们并不清楚具体有多少支球队削减了球探预算。多伦多蓝鸟队就是其中一支，他们也为此付出了代价，球队在 2002~2005 年球员选秀的结果都不尽如人意。但缩减预算绝不是当时的球队总经理、比恩的弟子J·P·里恰尔迪一时兴起的冲动做法，而是由于多伦多蓝鸟队的母公司罗杰斯通信公司正受加元贬值的困扰，不得已而为之。

《点球成金》已经出版 10 年了，球探和数据怪才之间的争斗使他们耗尽了力气。2004 年，红袜队在时隔 86 年之后再次捧得世界联赛冠军，成功的原因在于球队既充分尊重了球探的意见，也对已有数据进行了细致分析，两相结合成为两个阵营和平共处的关键所在。2003 年时还被界定为"球探"阵营的许多球队，如圣路易斯红雀队，此后也越来越多地运用数据分析的方法，这些球队现已成为最具创新精神的棒球队伍。而那些原本就偏好数据的"怪才"队，如奥克兰竞技队，非但没有削减球探预算，反而增加了相关投入。

2007~2009 年的经济衰退促使球队更频繁地采用数据分析的方法。尽管棒球界算是平稳地度过了经济危机，但仿佛每支球队都突然间变成了"点球成金"的球队，都需要将有限预算的价值发挥到极致。廉价的数

据怪才一直都供大于求：从哈佛大学和耶鲁大学毕业的经济与计算机专业的高材生，本可以在投资银行谋一份年薪达 40 万美元的工作，但他们却更愿意搬到坦帕或克利夫兰，日夜不停地工作，拿到的薪水却只有银行薪酬的 1/10。对于球队而言，花 4 万美元聘用一个数据怪才远比投入 4 000 万美元引进一个自由球员更值得，毕竟这位自由球员的技术水平注定会下滑，很快就会表现平平。

但这并不意味着数据怪才已经大获全胜了，如果说他们已经证明了自己的价值，那么球探同样做到了这一点。

系统与球探的对决：球探赢了

PECOTA 是“投手经验比对与优化测试算法”（Pitcher Empirical Comparison and Optimization Test Algorithm）英语单词首字母的缩写，这个冗长名称的缩略词正好是 20 世纪 80 年代堪萨斯城皇家队的内野手比尔·裴克达（Bill Pecota）的名字，那时的他对我钟爱的底特律老虎队而言一直是一个威胁。

PECOTA 预测系统最初是为预测投手——而非击球手——的表现而设计的。投手的表现极难预测，因此在用 WFG 系统进行过数年实验之后——你肯定能猜出来这个缩略词代表的含义[1]——《棒球规程》最终还是放弃使用它，于是年刊中的预测栏曾一度空白。意识到这是一个商机后，我向赫卡贝推荐了 PECOTA 系统。让我感到有些意外的是，他和《棒球规程》的同事们对这一系统一致认可，他们愿意以股权交换的方式购买 PECOTA 系统，条件是我得再开发一个相似的系统，用于预测击球手的表现。我欣然地接受了这笔交易。于是，第一组 PECOTA 预测结果便发表在 2003 年冬季的《棒球规程》上。

2003 年赛季结束时我们发现，PECOTA 系统确实比其他商业预测系统的效果要好一些。事实上，2003~2008 年这几年，无论在我们自己还是别人的测试中，PECOTA 系统的预测准确度总是与其他系统相当或高于其他系统，同时还推翻了拉斯韦加斯赌场惯用的台词。一些意外的成功使 PECOTA 名声大振，比如，芝加哥白袜队在 2005 年加冕世界联赛冠军，但 2007 年赛季 PECOTA 却预测该队只能获得 72 场胜利，惨淡结束赛程。这一预测必然招致芝加哥媒体和白袜队决策人员的一片骂声。但是，这个预测结果却惊人地准确：白袜队最终以 72 场胜利、90 场失利的成绩结束该赛季。

然而，在 2009 年前后，PECOTA 预测系统却被其他系统追上甚至赶超了。正如当年我借鉴詹姆斯和赫卡贝系统的优点一样，其他系统的研发人员也汲取了 PECOTA 系统的长处，并创建出自己的预测系统，而

且有些系统非常棒。各个系统都会对大联盟球员的表现进行预测，如果每年都对这些系统的最佳预测结果做个排序，你会发现，那些最棒的预测系统彼此之间的差距不过一两个百分点。

但我当时之所以创立 PECOTA，还有另外一个目的：对那些像佩德罗亚一样的小联盟球员的表现进行预测，这项任务可能要困难许多。因为当时鲜有预测系统对小联盟进行预测，直到最近才出现类似的系统，所以当时 PECOTA 系统唯一的竞争对手就是球探。

2006 年，我首次公布了 PECOTA 系统预测出的 100 名最具前景的球员名单，和同时期《美国棒球》刊登的球探所列的名单形成对比。PECOTA 系统依据这些球员入选大联盟后的前 6 个赛季中为球队做出的贡献对他们进行排名。

2011 年赛季，也是预测结果发布的第 6 个年头，此时的我终于可以打开“时间胶囊”，看看这些预测是否准确。虽然名单中的球员如今都还相当年轻，但他们中谁是明星球员，谁是替补球员，谁已江郎才尽，已经一目了然。

在最具发展前景的棒球球员名单中，佩德罗亚排在第 4 位，这是 PECOTA 系统的成功案例之一。当然，PECOTA 的成功案例可不止这一桩，该系统当年对伊恩·金德尔的职业前景也很看好，而《美国棒球》并没有把他列入名单。事实上，金德尔自 2006 年以来两次入选全明星队，还成了得克萨斯流浪者队的主攻手之一。PECOTA 系统还看准了麦特·坎普，道奇队的超级明星，2011 年麦特·坎普曾十分遗憾地与棒球“三冠王”失之交臂。PECOTA 系统当年的这个判断也比《美国棒球》高明。

上面几位也许你还有所耳闻，但你听说过乔尔·古兹曼、唐纳德·莫菲，还有约塞米蒂·佩蒂特这些人吗？若非资深棒球迷，很可能根本不知道这些人。但 PECOTA 系统也将他们列入了那份名单。

《美国棒球》难免会看走眼：球探们对布兰登·伍德、拉斯汀·米利奇以及马克·罗杰斯的职业前景十分看好，但他们此后的表现不尽如人意。球探们还认定红袜队的明星投手乔恩·李斯特、洛杉矶道奇队的游击手特洛伊·图洛维茨基以及巴尔的摩金莺队的外野手尼克·马卡提斯会有不俗表现，但他们在小联盟时一直表现平平，PECOTA 系统并未将其列入名单。

现在，我们有充足的数据可以对两个预测系统进行统计学分析。具体来说，我们可以查看每份名单上的球员为自己所在的大联盟球队带来的胜场数，这些胜场数以 WARP（超过替补球员的胜场数）的数据形式呈现，这就意味着要收集每一位球员为整场棒球比赛贡献的价值，包括

击球、投球及防守。

PECOTA 系统所列名单里的球员在 2011 年大联盟赛季贡献的胜场数总计为 546 场（见图 3–3），而《美国棒球》所选球员的贡献更大，胜场数总计为 630 场。尽管球探们的判断有时会出现失误，但他们的工作还是具有非常大的价值的：他们预测的准确度比单纯靠数据分析进行的预测高出 15%。这个差距听起来也许不算大，但带来的收益可不小。每赢得一场胜利，棒球队都愿意向球探支付 400 万美元。这样算来，在过去的 6 个赛季中，那些球探预测名单上多出来的 84 次正确预测，会为他们带来总计达 3.36 亿美元的收益。

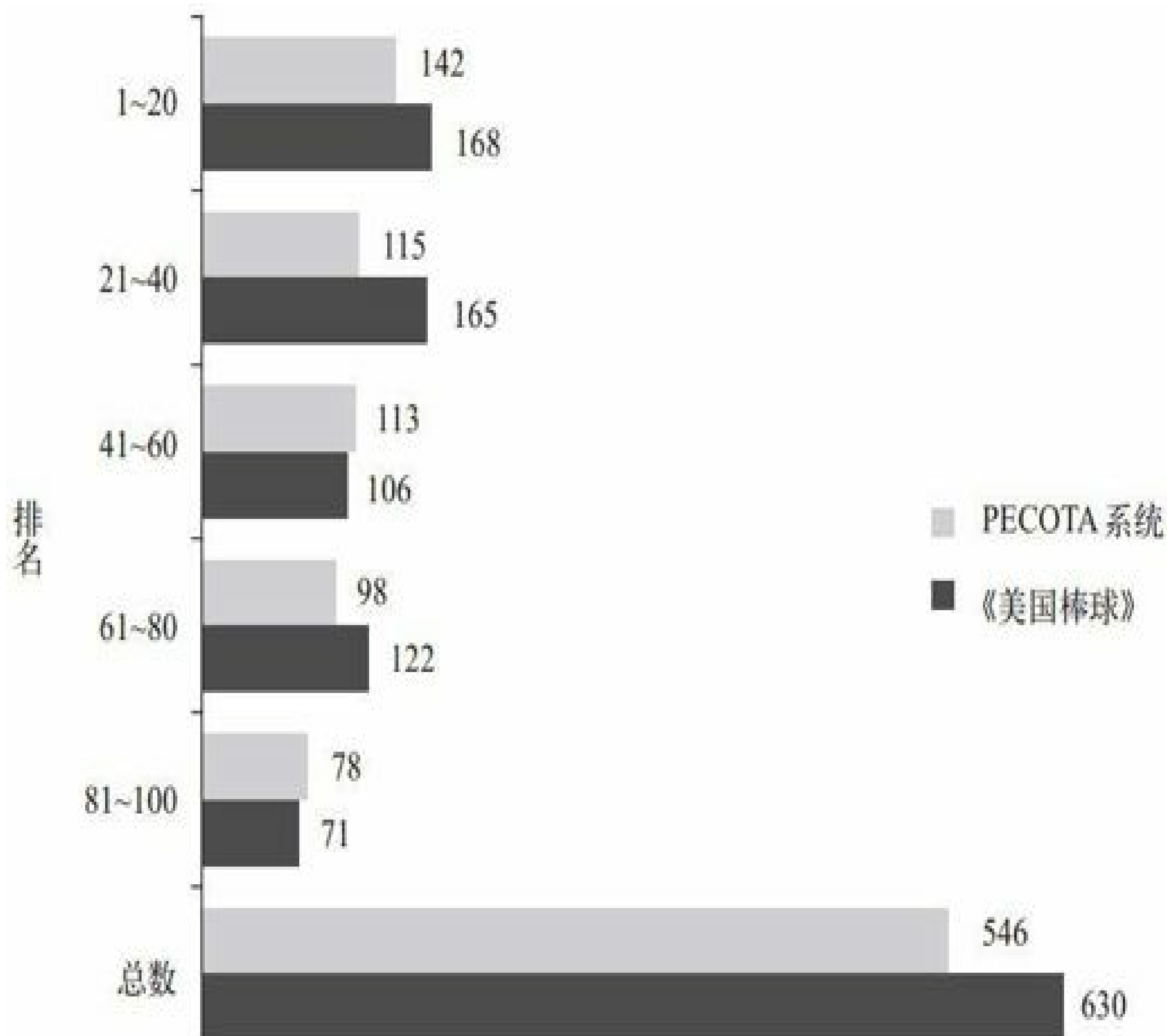


图 **3** 系统和《美国棒球》列出位最具前景球员年赛季贡献的胜场数

球探和数据怪才的偏见

如果 PECOTA 系统的预测结果能比球探们的预测结果更准确，那自

然很好，可我并不期待出现这样的结果。就像我在 PECOTA 预测名单公布后不久所写的那样：

> 这次球探对战数据怪才的戏码引发诸多趣闻，但我并不期待 PECOTA 系统给出的排名能与《美国棒球》的排名同样准确。
>
> 预测系统的工作原料是信息。能够同时将球探的推断和统计信息收入囊中，就意味着你拥有更多原料。单纯依靠数据的预测想要打败依靠综合分析的预测，只有一个途径，那就是综合分析的过程中掺杂了过多偏见，若偏见太强，就会削减掉这种分析手段的优势。

换句话说，那些球探用的是综合分析的方法，他们得到的信息并不局限于统计数据。球探和 PECOTA 系统都会观察球员的平均击球率和防御率，PECOTA 系统毫无主观偏好，也善于从数据中排除部分噪声，并将这些数据排列好。而球探则可以了解到很多信息，这是 PECOTA 系统做不到的。比如说，球探并非只能依据投手三振出局数推断球员的表现，他们还能够拿出雷达测速仪测定投手投出的快速球的速度，或者用秒表测定他的跑垒时间。

这一类型的信息更接近我们进行棒球比赛结果预测的根本目标。在美国棒球小联盟中，投手即使投出一个无力快速球也可得分，只要投到好球区，或几种投球混合使用，都可以让不少击球手三振出局，因为大多数击球手实力平平，很容易被战胜。但是在美国棒球大联盟中，击球手甚至能将时速将近 158 公里（98 英里）的快速球击出场外，此时投手只能依靠软抛球谋得胜算。结果是，PECOTA 系统很容易被这些伪正数欺骗，但是球探却能够辨别哪些球员具有大联盟水平，哪些球员还需要修炼。

当然，任何的人为判断都有可能掺杂个人偏见。就像本书前文中提到的那样，对预测持错误态度的人，常以信息过多为借口提出一个不公正的理论，用来描绘世界应该是什么样的，而不是体现客观事实，故而他们获得的信息越多，情况就越糟。

也许在前“点球成金”时代，球探们带有更多偏见。他们倾向于从美学的角度评判球员，比如球员的球衣合不合身，而不是看这个球员是否具有天赋。若说近些年来位列《美国棒球》预测名单中的那些球员都有不俗表现，那么 20 世纪 90 年代，该刊力荐的那些球员的表现就平平了，像托德·范·波佩尔、鲁本·里维拉和奥布莱恩·泰勒等球员的职业前

景明显被夸大了。

数据怪才们也有自己的偏见，其中最致命的一个偏见：若某项因素很难量化，那它就被视为无关紧要。例如，在棒球运动中，很长时期内防守质量要比击球率或投球质量更难测量。20 世纪 90 年代中期，比恩率领的奥克兰竞技队忽视防守，常让马特·斯特尔这类庞大笨拙的球员防守外野，而马特·斯特尔是一个天生的击球好手，却总被安排去防守外野。随着防守技术研究的不断推进，奥克兰竞技队无可避免地为其薄弱的防线付出了巨大代价，每个赛季奥克兰竞技队得多输 8~10 场比赛。在这种情况下，不管球队的安打表现多好，都不可能跻身热门球队之列了。比恩从中吸取了教训，最近由他带领的球队防守得都不错。

在预测美国棒球小联盟球员的表现时，这些盲点带来的失误会造成更大的损失。对于一个颇有成就的美国棒球大联盟球员来说，关键是预测他能否续写辉煌。一个聪明的系统可能会预测，球员的发展走势会上下浮动几个百分点。但如果只是简单地假设这个球员下个赛季的表现与前几个赛季相仿，你也不会错得太离谱。通常情况下，球员日后的实力与现在的表现不会有天壤之别。

然而，在主要的职业运动中，棒球属于比较特殊的，因为这项运动中包含着极其复杂的小联盟系统。美国橄榄球联盟很少批准小联盟赛事，美国第一大职业篮球赛事的小联盟数量也不多，但在棒球界，每个大联盟下附有 8 个小联盟，共有 240 个小联盟。另外，足球和篮球队员可以从大学甚至高中直接进入职业联赛，变身为球星，这种情况在棒球界却很少见，即使是最有天赋的新秀，也必须在比利斯特、贝克斯菲尔德或宾厄姆顿苦苦地等待进入美国棒球大联盟的机会。

对这些棒球手的表现进行预测是非常具有挑战性的，因为我们一直在期待这些球员最终能做到他们现在无法做到的事情：参与高水平的赛事，成为大联盟球星。美国棒球大联盟选秀赛非常残酷，就连全美高中时期表现最棒的击球手布莱斯·哈珀，在面对大联盟（选秀）投球时都可能惨遭淘汰。要知道大联盟是勤奋和运气结合的地方，新秀们必须变得更强、更壮、更聪明、更守纪律，才能在大联盟打比赛。假设你来到一所普通高中的课堂上，你应该对那些学生留心地观察几天，然后预测他们当中谁将来会是医生、谁会是律师、谁会是企业家、谁又只能勉强度日。我猜你一定会参考这些学生的成绩和学术能力评估测试分数，再看看谁的人缘更好，但你还必须进行一些大胆的猜测才行。

很多业余球探（以及所有模仿球探的做法设计出来的数据预测系统）都是这么预测的。尽管一些棒球球员是从大学中选出的，还有很多球员是直接从高中选出的，但球探的“侦察”活动在这些球员十几岁时就

开始了。和同龄人一样，这些球员荷尔蒙爆发，经历着青春期的焦虑，身体正在发育，并且需要与酒色的诱惑抗争。如果你把你企业的未来交给这样一群 19 岁的年轻人，你能放心吗？

生理指标与心理指标

正如刘易斯在《点球成金》中提到的那样，有些球员拥有惊人的天赋却不自知，比利·比恩就属于这样的球员。1980 年第一轮选秀后，比恩在大联盟只打了 148 场比赛，职业生涯的平均击球率仅为 0.219。但与约翰·桑德斯相比，比恩的球员生涯还算辉煌，至少他还入选过一次名人堂。

桑德斯现为洛杉矶道奇队的球探，也曾参加过大联盟比赛，确切地说，他只打过一场比赛，就像《梦幻成真》里的“月光先生”葛兰曼博士一样。1965 年 4 月 13 日，在堪萨斯城竞技队对阵底特律老虎队的比赛中，19 岁的桑德斯作为替补跑垒员，在第七局时上场。但桑德斯当时毫无上垒准备：两个击球手突袭成功，他在下一局开始前就被换下场。从那之后，他再也没有参加过美国棒球大联盟比赛。

桑德斯并非缺乏天赋，他在内布拉斯加州格兰德岛高中读书时一直是位全能体育明星：1963 年在州足球队踢四分卫，1964 年效力于州篮球队，还曾经得过州田径运动会铁饼金牌。棒球可能并不是桑德斯最擅长的运动项目，但是他真的极具天赋，1964 年夏天高中毕业时，他就拿到了奥克兰运动家队职业球员的合约。

但新秀津贴规则阻碍了桑德斯的发展，断送了他的大好前程。1965 年，当时还没有采用大联盟合约，所有业余选手都是自由球员，球队可以根据球员意愿支付薪酬。为防止经济条件较好的球队笼络天资好的球员，新秀津贴规则提出了一项惩罚措施：每个得到大笔签约津贴的球员必须在大联盟球队打满两个职业赛季。但事实上，他们的实力根本就打不了这样的高水平比赛。

这项规则实际上是让像桑德斯这样前途光明的球员吃了“杀威棒”。大多数得到大笔新秀津贴的球员只能坐冷板凳，几乎没有机会去打任何大联盟的比赛。在最需要每天都打比赛积累经验的时候，这些年轻的球员却和比赛无缘。球队的支持者和队友也在纳闷，球队为什么花重金雇这些乳臭未干的 19 岁“娃娃”来当球童？所以，这些人也不可能理解和同情那些新的年轻球员的困境。在这种情况下，除了桑迪·考法克斯和哈蒙·吉尔布鲁这样的得到签约津贴的球员能咬牙挺下去并最终入选名人堂，当时很多具有潜力的优秀球员在这样困境中都撑不下去。

从内布拉斯加州最佳运动员到《棒球百科全书》的一个小脚注，这

样高低起伏的人生经历让桑德斯对年轻运动员的心智有着独到见解。一天上午，我通过电话联系到桑德斯，当时他正从北卡罗来纳州开车前往芝加哥观看勇士队的最佳附属队的比赛。

早在 2003 年新奥尔良冬季会议之际，我就觉得桑德斯会孜孜不倦地为棒球付出一生。（短暂的）运动生涯结束之后，桑德斯还在为棒球运动贡献力量。但桑德斯从未从“球探对战数据怪才”的角度看待棒球运动。

桑德斯告诉我说：“我喜欢估算，喜欢统计学证据，就连从前使用计算器或其他计算仪器处理数据时都觉得很享受。”他接着还说了一件逸事：“一个球探曾说过，‘伙计们，面对现实吧，每天我们到球场后做的第一件事是什么？是到记者席拿最新的数据’。我们拿的是数据啊！那样做有什么不对吗？可你们就是那样做的。”

的确，自棒球运动诞生之日起，统计数据就一直是其中的一部分。第一份新闻技术统计数据——其中包括得分、安打数、出局数、助杀数及失误次数——是由亨利·查德威克于 1859 年发表的，而第一个职业联盟直到 12 年后的 1871 年才成立。到了“点球成金”时代，许多争论的焦点不是该不该使用数据，而是该使用哪些数据。比如詹姆斯等分析家多年前就提出，和安打率相比，上垒率和得分的关联更大，但这一观点一直不被业界的传统主义者所认可。

这样的争论通常涉及那些数据怪才的专业领域，可谓投其所好了。上垒率比平均击球率更有用，投手防御率比他的输赢纪录更能客观体现其水平，这些都是科学事实，就像“地球围绕太阳旋转”一样不容反驳，数据怪才的这些看法自然没有错。对此类争论，数据怪才们只赢不输，这也使得他们有些自鸣得意，忘了自己在别的问题上其实很含糊。

离美国棒球大联盟越远，统计数据的用处就越小，因为此时你需要做得更多的是对球员的表现进行预测，而不是测量。对于那些级别比较高的小联盟，比如 2A 等级或 3A 等级小联盟，统计数据的预测作用和大联盟的几乎无异。但是级别较低的小联盟数据就没有那么可靠，大学或高中球员的数据的预测力则更弱。

球探们（预测时）经常会用另外“五大指标”来替代统计数据：击球效力、击球平均数、速度、手臂力量以及防守范围。但这些指标饱受诟病，也确有缺陷。这 5 项指标并未包括送出保送球和避免三振出局这样的垒上规则。虽然这 5 项指标会让人觉得它们同等重要，但实际上，除了游击手和捕手之外，对其余所有位置上的球员而言，击球效力要比手臂力量重要得多。

“五大指标”之所以饱受批评，还有另外一个原因，那就是仅凭这几

个指标获得的信息并不全面。随着一名小联盟球员的球技日渐成熟，他的统计数据应当越来越多地体现出这几项指标——当然，很可能这位球员的水平并没有多少提高。实际上，其中有些指标就是统计数据："击球平均数"表现为平均击球率，"击球效力"表现为二垒安打数和本垒打数。如果球探告诉你某位球员击球效力得了 70 分（满分 80 分），但这位球员在阿尔图纳弧线棒球队效力时，每年却只贡献 10 个本垒打，那么你觉得球探给出的这份报告可信度高吗？

作为业界资深人士，桑德斯对过分强调"五大指标"的做法持怀疑态度。他告诉我说："这些指标也没有多么玄妙，无非是快速奔跑、用力投球之类的技术，其实人人都懂，只是球探可以在第一时间看到这些数据。问题是，这些技术真的可以有效地塑造得分球员吗？那些指标真能转变为可用的技术吗？以击球速度为例，若一个球员击球的速度很快，但他却不相信自己，总是想去做投手，那他的击球速度就成了不可用的技术，也就是说他的击球效力没有转变为可用的技术。"

桑德斯不太注重这些生理指标，而是更加看重可用的、适用于比赛的技术。桑德斯认为球员的心理指标可以决定其生理转变的程度。心理指标通常比生理指标显露得晚些。桑德斯的妻子是一位特殊教育工作者，在她的指点下，桑德斯作了一项研究。这项研究表明，大多数人在 24 岁之前都处于心理未成年期。对于那些年龄小于 24 岁的球员，如果桑德斯发现其各项心理指标正在发育的话，就会暂缓将其从名单中除名；对于那些年龄大于 24 岁的球员，就要看其球场表现了。有趣的是，24 岁正好是球员在 2A 级小联盟中效力的年龄，此时刚好可以开始根据其统计数据来预测其球场表现。

桑德斯并没有明确地指出一个球员的心理指标应当包含哪些内容，但是通过我们之间的交流，对于他认为有助于预测大联盟胜率的五大智力和心理能力，我进行了总结。

赛前准备和职业道德。与其他职业运动不同，棒球队员一周有六七场比赛。足球或篮球队员只要比赛当天进入状态即可，但是棒球队员不行，他们每天都要保持参加职业赛事的状态。这就意味着他们必须进行一定量的科目训练。桑德斯喜欢提前到达球场，因为他总觉得赛前预热（相比比赛过程）更能体现出队员的科目训练效果。例如，前文中提到的那个 9 月的晚上，在芬威球场进行场内练习时，佩德罗亚显然比队友更专注。他在进行常规训练，无法忍受任何干扰，这干扰当然也包括某个不知名的记者试图对他进行采访。

全神贯注和集中精神。尽管和赛前准备有关，但这一项特指球

员在比赛过程中自我控制的习惯。棒球是条件反射型运动，击球手只有 3/10 秒的时间决定是否挥棒击球，内野手在地滚球离垒的瞬间就要做出反应。桑德斯说："球员若精神不集中，我们也拿他们没辙，但我希望游击手和中场内野手能集中精神应对每次投球。"

竞争力和自信心。所有职业运动员天生就具有竞争力，这种说法似乎成了公认的道理，所以棒球手在职业生涯早期必须克服自我怀疑和其他心理障碍。前一秒还是高中的风云人物，下一刻却在安纳波利斯到格林斯博罗的大巴车上读着网上有关自己比赛失利的报道，心情立刻跌到谷底，这些都得学着自己调节。每当桑德斯看到一个颇有天赋的球员难以打出成绩时，他总在想：有没有一种成功的欲望能与失败的机制抗衡？对成功的渴望能否足以战胜对失败的恐惧？

压力管理和自谦。棒球比赛中，即使技术最棒的击球手多数时间也得不到分，每个球员在赛季的某些时段都会进入低谷状态。应对这样的挫败需要差记性和幽默感。桑德斯最喜欢的一个侦察策略，就是观察球员在碰到一场硬战或运气不好的比赛时的反应。"击球手大力挥棒，动作夸张却没打中球，球迷会觉得很滑稽，这时，我喜欢俯身看到他的微笑。这就意味着，接下来一击——嘣！——球飞出 0.12 公里之外。"一旦球员进入美国棒球大联盟，就必须接受球迷和媒体的品头论足，此时这项心理技能就显得更加关键了。

灵活应变和学习能力。比赛过程中，球员处理突发情况的能力如何？是否听从教练指挥？生活境遇发生改变时如何调节？转会或是被安排到新的位置时该怎么办？要知道，即使是天赋极高的球员，从业余队员到专业队员的转变也绝非一帆风顺，因此，一个出色球员处理问题的手段不能太过僵化。桑德斯评论说："沿着一条走廊走到拐角，有些人会从容地绕过去，而不是猛烈地拐过去。能玩转比赛的人就是这样的聪明人，他们知道如何克制自己的紧张感。"

当然，上述几种能力在很多人类活动中都颇为重要。其中一些已经为预测者所用，尤其是桑德斯所说的灵活应变能力：遇到新信息时，你会怎么处理？情况发生改变，证据指向其他方向，此时反应过激或无动于衷都会导致预测失败。

然而，就竞争的激烈程度而言，几乎没有哪种职业运动能与棒球匹敌。数十万名业余球员，数千名职业球员，只有 750 人能在指定时间加入美国棒球大联盟，而这 750 人中能入选全明星队的则更是少之又少，只有几十人。桑德斯的工作就是寻找那些不轻易言败的明星球员，他自己也像那类球员一样恪尽职守、努力工作，虽已年近古稀，几乎每天都

坚持出门工作。

桑德斯确实为道奇队提供了最具价值的信息，别人提供不了这类信息。

信息是决定预测成败的关键

《点球成金》的主人公比利·比恩认为不断地收集信息是成为好球探的秘诀。

“如何定义好球探？球探可以找到别人找不到的信息。要了解一个人，还要了解这个人的孩子的信息，了解他的家庭信息等。很多信息都得你亲自去找。”他如是说。

比恩应该清楚，奥克兰竞技队的大多数胜利有赖于该队出色的数据统计能力，但球探对业余选手的甄选工作也为球队的优异成绩做出了重要贡献。21 世纪初载入《点球成金》的该队球星中，米格尔·特哈达、杰森·吉昂比、巴里·奇托、提姆·哈德森、艾瑞克·查韦斯等多数都是由俱乐部签约并培养的。

比恩告诉我，奥克兰竞技队十分着迷于数据型综合分析，因此该队如今的球探预算比以往都高。如前文所说，棒球球员打满 6 个赛季之后才是自由球员，那时他们至少有 30 岁了，按照比尔·詹姆斯的老化曲线来看，那些俱乐部花在自由球员身上的钱通常无法得到应得的回报，毕竟球员的最佳状态已经基本过去了。但从另一方面看，球员在 30 岁之前还可以为球队做出巨大贡献。而且，在棒球经济学中，比较年轻的球员通常都是低价“买入”的。

如果和其他行业一样，从盈亏角度看待棒球，这个行业中几乎所有的价值都是由球探甄选和培养球员的过程创造的。如果一支球队的预测系统特别完善，也许该支球队只需支付 1 000 万美元的年薪就能获得一位价值 1 200 万美元的球员。但如果这支球队的球探团队足够出色，也许只需花 40 万美元就能签到那位价值 1 200 万美元的球员。在奥克兰这样的小市场中只能智取。

所以，奥克兰竞技队从来不会对球探表示不敬，而是非常尊重这些人。比恩还明确表示，球队纳新时一定会考量球员的心理素质。球队仍十分信任缜密的分析，球队在处理收集的信息时，秉持着严谨的作风和纪律，但并不随意排斥某种信息。

“每个球队赋予客观分析和主观分析的权重都不同。”比恩解释道：“以我们奥克兰队的立场来看，从某种意义上讲，我们只能做出客观判断，不能依靠直觉。若某次直觉判断碰巧是正确的，我猜这只是歪打正着而已。我们的工作职责决定了我们不能乱下决定，然后企盼好

运。如果我们在玩 21 点，庄家底牌有一张 4，我们手里有一张 6，遇到 16 点，我们就爆牌了。”

如前文所述，预测成功的关键是，不应该局限于定量信息，而应该用心权衡信息的适用性。比恩的工作宗旨是尽可能地收集信息，尽可能严谨、专业地分析信息。

信息越多，你的预测就越准确，如果是这样，那你就是一个十分出色的预测者。若态度不端、习惯不良，那你就会像菲利普·特罗克的政治学者一样，预测得十分糟糕。候选球员 A 的平均击球率为 0.300 外加 20 个本垒打，轮休日出现在“救济厨房”帮忙，球员 B 也有相同的击球率和本垒打个数，但平日里他逛夜店、吸毒品，两位球员的区别尽管很难用数字量化，但是你进行预测时一定会考虑这个区别。

事实上，很多时候定性信息有可能转变为定量信息。球探实际上是用严格的数据范围来测评球员，每类数据的范围都是 20~80 之间，当然你可以将这种方式归入数据模式中，和球员平均击球率放在一起，看看其价值何在。有些球队，如圣路易斯红雀队，已经开始尝试这一模式了。

确实，在棒球界，球探模式和统计数据之间、定性信息和定量信息之间的差别已经非常模糊了。例如，投手球路分析系统（Pitch f/x）是一套三维相机系统，如今在大联盟中的每个球场都已经安装了。投球数据折线图不仅可以测量球速——很多年前雷达测速仪就能做到这一点——它还可以测量出球落垒前在垂直方向和水平方向的移动距离。比如，我们现在可以用统计学方法来分析扎克·葛兰基，他是 2009 年大联盟最佳投手，是赛扬奖获得者，是密尔瓦基酿酒人队的队员，他还投出过最漂亮的弧线球。我们还可以用另一种方法来分析，马里亚诺·里维拉的卡特投球果然名不虚传。从传统角度来讲，这些信息都属于球探的考察范围，现在成了预测系统中的变量之一。

不久的将来，我们也许就能用一套完整的三维成像系统记录球场上发生的一切了。雅各比·艾尔斯布里一跃而起接住头顶飞来的球时，我们可以立刻测量出他这一跃有多高；也可知道铃木一郎完成一圈跑垒的确切时间；或是当雅迪尔·莫力纳想将对方盗垒手赶下垒时，在二垒奋力接球时的速度究竟有多快。

正如当年《点球成金》中公布的预测结果一样，这项新技术并不会减少球探的数量，但它有可能会被重点用来测量更难量化、更加隐秘的信息，例如球员的心理指标。像桑德斯这样聪明的球探在这一方面已先行一步了。

并不是信息越多，预测就越成功

为什么球探们对达斯汀·佩德罗亚的预测错得那么离谱？

所有的球探对佩德罗亚的基本信息都没有疑义，他们都知道佩德罗亚是一个非常出色的击球手，垒上技术灵活，心理素质超群。球探们也都知道佩德罗亚挥棒姿势怪异，防守技术稳定但不突出，移动速度平平，且个子不高，体格一般。

然而，作为年轻球员，佩德罗亚的个人履历很特殊，很多球探不知对此该如何看待。桑德斯告诉我："招募新球员时，球探都是带着主观想法进行选拔的，他们心中早就有了典型标准。佩德罗亚在很多方面都不符合这些标准，他需要新的标尺。"

当我们无法把一个方形物体塞进圆洞时，通常会责怪这个方形的物体。有时，就是思维定势使我们没有及时做出调整去适应新情况。我们总是本能地将信息归为不同类别，通常类别的数量相对较少，这样便于跟踪。（美国人口调查局将数百个种族团体分成六大种族类别，数千名艺术家按照音乐类型归为几类，这些都是归类的做法。）

在大多数情况下，归类的方法还是非常奏效的。可是，一旦遇到不易分类的信息，我们通常就会无视它或对其进行错误分类。这也是比恩不愿"凭直觉"作决定的原因之一。如果过分依赖第一印象，很可能会错过潜力无穷的球员，而奥克兰竞技队的经济条件可负担不起这样的过失。

像 PECOTA 这样的预测系统是在数千名球员中搜索信息相似的球员，对球员进行分类时更需要谨慎细致。PECOTA 系统能够将佩德罗亚的各项技术归入更适合的类别之下，也确实搜索到了一些与他的信息相似的前辈。

若将佩德罗亚的其他技能考虑在内，他矮小的身材反倒可能是他的优势。棒球的好球区是指从球员肩膀到膝盖的范围，个头越矮，投手的目标区就越窄，而佩德罗亚对来球的捕捉能力又强，这样他就能更好地利用自己个子矮、好球区小的优势。

另外，二垒手身材矮小，防守起来更有优势。二垒手要求灵活敏捷，能像猫一样灵活地捕捉地滚球。史上最佳的二垒手大部分都是矮个子。入选名人堂的 17 名二垒手中，只有内普·拉如瓦和瑞纳·桑德伯格身高超过 1.82 米。最伟大的二垒手乔伊·摩根的身高只有 1.70 米。

那些球探非常擅长本职工作，也善于分类，但是这一次他们预测得过于仓促，偏见也过大。佩德罗亚的矮小身材从某些方面来看其实是优点。

但是，一切都没有定数：PECOTA 系统并没有认为佩德罗亚一定会成功，只是觉得可能性较大，而球探并不这样认为。不过，红袜队信任佩德罗亚，这才是最重要的，而佩德罗亚的自信也给球队带来了好运。

第一次遇见比尔·詹姆斯是在 2009 年 10 月“纽约客艺术节”的一个专家小组。在随后的一个奇特的派对，比尔·詹姆斯穿了一件色彩夸张的运动衫，脚踩一双又大又旧的木屐，在一群打扮时髦的人中，显得尤为特别。派对上其他人都在追随明星苏珊·萨兰登，我和詹姆斯却在吧台边聊天。

詹姆斯在红袜队的职责范围很广，也很秘密（很多细节他不方便透露）。过去的 25 年里，作为一名在场外摇旗呐喊的人，詹姆斯写了很多关于棒球的文章，人到暮年他变得平和很多。现在，作为一名熟悉行情的内行人，这项运动对詹姆斯有了不同的意义。詹姆斯是最早从心理角度审视棒球的人。

詹姆斯对我说：“年轻时我写的东西，很多观点都不对。自从有了孩子，我的态度发生了很大的转变。这句话是有些老生常谈，但确实，有了孩子才开始懂得每个人都是父母的宝。这就是内行人和外行人的区别。小时候觉得这些人是电视里、游戏中或是棒球卡上的人物，并没有意识到他们也是普通人，正在尽全力比赛。”

令我惊讶的是，尽管詹姆斯、比恩和桑德斯这三位看待棒球这项运动的角度存在极大差异，但他们对很多问题的评论却非常相似。事实上，如果把我和他们 3 位的交谈记录放在一起，你很难指出哪一段话是谁说的。（詹姆斯的话相对容易分辨，因为他更风趣）。詹姆斯逐渐注意到球探为红袜队带来的价值，他认为这也是他的使命。在棒球运动中，衡量预测是否准确的方式很严苛，只看比赛的输赢，所以，所有的预测者都明白预测不能离谱。如果信息越多，你的预测却越离谱，那你注定会失业，也没有资格得到麦克劳林媒体集团的终生聘用合约。

“从某个角度看，我和其他球探看待棒球的方式是非常相似的。”詹姆斯接着说，“在政治领域，从最‘右翼’到最‘左翼’，人们说的其实完全是一回事。也许球探行业也是这样，别的球探努力探寻的问题和我努力探寻的问题完全是一回事儿。”

2004 年，詹姆斯协助红袜队进行职业棒球选秀，佩德罗亚是第 65 个被选中的。詹姆斯写了一份报告，其中对佩德罗亚赞许有加，但还是建议球队选择另外一个人。虽然詹姆斯对这个选择很满意，却也很开心地看到佩德罗亚后来用行动证明詹姆斯当时的评判很愚蠢。

然而，在佩德罗亚职业生涯的早期，连他的一些头号粉丝有些时候都会怀疑他。2006 年 8 月，佩德罗亚开始了自己的第一个大联盟赛

季，打了 31 场比赛，平均击球率仅为 0.198，只有 6 个二垒以上的长打。没人特别关注他的表现；更令人感到意外的是，在那一季的最后几周里，红袜队无缘季后赛已成定局。于是，新英格兰地区的焦点完全转移到卡尔特人和爱国者这两支球队上。第二年，佩德罗亚担任球队的专职二垒手，但他进入状态还是很慢，赛季已经开始一个月了，佩德罗亚的平均击球率只有 0.172。

如果当时他是在芝加哥小熊队这样的球队，佩德罗亚很可能会被开除。小熊队的决策机制从前是出了名的随意，直到近几年才有所改观。对很多球队来说，每个行动都会遇到同样的过激反应，但是红袜队不会，这支球队是经过系统方法训练出来的。詹姆斯透露，从佩德罗亚在那个赛季的表现中，红袜队看到了令人欣喜的地方。佩德罗亚一直在积极应对比赛，制造了很多接球机会，只是球没有落在球棒上而已。照这种情形看，比赛的各项数据极有可能会向他倾斜。

詹姆斯对我说道："我们都有对数据失去信心的时候，或许你也清楚这一点。回看过去一年的情况，当时佩德罗亚的击球率仅为 0.180，再看看他的挥杆失误率，可能高达 8%~9%。在那个春季赛季中，佩德罗亚一直在努力，但他同样也对自己的数据失去了信心。按理说，那么卖力地挥杆，接球机会不可能那么少，击球率不可能只有 0.180。"

红袜队当初选择佩德罗亚并非轻率之举。他们当时一直都在留意观察佩德罗亚的比赛表现，是"雪藏"佩德罗亚还是派其首发，球队都要慎重思考。球队会在更大的背景下考虑这个问题，不会让数据左右他们的决定。

詹姆斯告诉我，红袜队唯一担忧的是佩德罗亚是否开始怀疑自己的实力。遇到类似情况，其他球员也许会对自己产生怀疑，但佩德罗亚不会这样，骗子和批评家都奈何不了他。

"幸好佩德罗亚是一个骄傲的家伙。如果他胆小怯懦，容易受到批评的影响，他就完蛋了。佩德罗亚并不理会旁人，依然我行我素，保持怪异的挥杆姿势，最终实现了逆转。"

佩德罗亚拥有桑德斯所说的"大联盟记忆力"——忘得快。佩德罗亚不会为糟糕的表现所困，因为他完全相信自己的打法没有问题，从长远来看，这一点真的很重要。但对所有让他分心的事情，佩德罗亚则采取零容忍的态度，这种态度无法将他塑造成最宽容的人，但这却是红袜队二垒手最需要的品质，也是他唯一在乎的事。

"缺点和优点是一对亲密伙伴。佩德罗亚善于化腐朽为神奇。"詹姆斯点评道。

《点球成金》的真正意义

在与比利·比恩讨论《点球成金》时，比恩说："正如迈克尔·刘易斯所说，争论已经结束了。"《点球成金》这本书有段时间曾对球探形成威胁，貌似球探的工作和生活也因此出现了危机。但这只是臆想，电脑从来没有取代球探。实际上，无论是由球探报告指导，还是靠数据系统分析，了解棒球市场未来所需的不同类型球员一直是供不应求的工作。球队挑选谁、与谁交易、付给自由球员多少钱，这些都决定着世界联赛的结果和数百万美元的走向。现在，当球队做决定时，会使用越来越多可支配的工具。在别的领域，信息革命并没有任何益处，但在棒球行业，信息革命带来的价值配得上那张数百万美元的高额支票。因为体育是高科技、激烈竞争、均衡的激励制度和大量数据的独特结合体。

但科技并未使比恩的生活轻松多少，其他球队已经复制了奥克兰竞技队的成功诀窍，为此他表示出自己的担忧。比方说，现在几乎所有球队都懂得上垒率的重要性，也都十分重视防守的作用，而没有发生改变的是，那些球队仍然（像从前一样）比竞技队富有。

在像体育运动这样竞争最为激烈的行业中，最出色的预测者必须不断地创新才行。树立"挖掘市场盲点"的目标容易，但这个目标并不能真正地为你谋划如何找到市场盲点，也无法确定这些盲点是代表希望的曙光还是错误的导向。拥有别人想不到的想法很难，拥有好的想法更难——即便有了好的想法，也很快会被人复制。

正因为如此，本书避免推崇速效方案，这些方案通常会让你以为，只要稍稍改变经营方式即可提高竞争力。优秀的创新者通常既从大局出发，又考虑细节。有时你可以从一个问题的最微小的细节中发现新的想法，但别人却不愿意这样做。当你进行最抽象的哲理思考时，会琢磨世界为什么是这个样子，想知道是否还会有另外一种主流范例可供选择，此时也可能有新的想法迸发出来。但大多数时候，我们都生活在宏观与细节中间的"舒服地带"，这种状态下很少能萌发出新的想法。在常态生活中，我们所作的分类和估值通常能行得通，但有时还是会漏掉那些有助于提高竞争力的信息。

要在正确的地方寻找信息和想法，关键在于要开发一些手段、培养一些习惯，这样才能在准确的地方发现更多的想法和信息。一旦发现了一些必需的技能，就要把它们变成能分输赢的技能。

这并非易事。但是，棒球会为创新者提供一个特殊的验证平台。10年前 PECOTA 系统首次亮相后，就几乎再也没有突破性的预测系统问世了，但一定会有人巧妙地利用 Pitch f/x 的图像数据，或是想出将球员

表现的定性数据和定量数据相结合的方法。很快这些假设都会实现，或许本书付印时就会实现。

“如今，进入棒球领域、创造领域及智力领域的人都十分优秀，10 年后如果我再去应聘球探这份工作，也许连面试的机会都没有。”比恩自嘲道。

《点球成金》已经落幕。

《点球成金》的影响极为深远。

[1]服务请求转发软件的发布将为你解决这一切烦恼。利用此软件你只需进行简单的设置就可以很好地将实际的服务器保护起来，并且不影响你的服务器的访问情况进行各种统计

第四章

天气预测：蝴蝶扇动翅膀，有可能引起龙卷风

2005 年 8 月 23 日，星期二，一架空军侦察机在巴哈马群岛上空飞行时，捕捉到一些干扰气流，几个旋涡正呈顺时针方向形成一股旋风，并由东向西移动，从大西洋逐渐向美国推进。从云层或卫星数据中很难检测到中断的风力模式，但海面上的货船已经有所察觉。美国国家飓风中心认为，足以把此次干扰气流描述为热带气旋，并将其标记为第 12 号热带低气压。在大西洋流域的所有热带气压中，有一半最终会演变为飓风。这次风暴很诡异，可能会加剧，也可能会消失。

这次的热带气压很快就加强了，截至星期三下午，美国国家飓风中心的计算机模拟机已经预测到有十几场飓风会在美国登陆，第一场将覆盖佛罗里达州南部，第二场可能会将气旋“带到”新奥尔良。风暴的威力已经足以升级为飓风，人们将它命名为“卡特里娜”飓风。

卡特里娜飓风第一次登陆时，持续的时间不长，它只是从迈阿密北部经过，几个小时后，又掠过佛罗里达大沼泽地，当时只算得上是 1 级飓风，不足以对太多人畜造成威胁。如果说那时卡特里娜飓风还没有从风暴中汲取太多力量，那么当卡特里娜飓风到达墨西哥湾的温暖水域时，就开始积攒力量了。星期六清晨，情况开始变得严重：卡特里娜飓风风力已达到 3 级，并且正在增强至 5 级。预测跟踪显示，它正从佛罗里达的狭长地带逐渐向西移至密西西比和路易斯安那。计算机模拟机已得出一致结论：新奥尔良必将遭到此次风暴的袭击。

我请美国国家飓风中心主管马克斯·梅菲尔德回忆一下他从什么时候开始意识到这次风暴的严重性。马克斯·梅菲尔德说：“在卡特里娜飓风来袭后，我已经参加过 5 次国会听证会了。当时有人问我是从什么时候开始关注新奥尔良的，我告诉他是 60 年前。”

新奥尔良遭到严重飓风的正面袭击对于每个气象预测者来说，在很长一段时间内都是一个噩梦。这个城市的各方面情况都适合自我破坏甚至灭亡。这一方面与那里的地形有关：新奥尔良几乎是陷在墨西哥湾中，很大一部分人口生活在海平面以下，他们的保护措施只有过时的堤坝和一些几乎已经被冲进大海的天然栅栏。另一方面与当地的风俗有关：新奥尔良人在许多方面做得相当好，唯有两件事他们不屑去做，那就是他们不愿迅速行事，也不信任权威，这就是新奥尔良人。如果他们不这样的话，那新奥尔良也称不上是新奥尔良了。这座城市本可以更好

地应对卡特里娜飓风，因为这两方面的事正是飓风来袭时最应该做的事。

美国国家飓风中心对卡特里娜飓风的预测是准确的，在堤坝被冲毁前 5 天就预测到这次袭击有可能会出现，在飓风登陆前 48 个小时就看出这次噩梦的一些端倪。20~30 年前，根本不会有这种先进的预警，人们也无法及时撤离。在过去几十年中，美国国家飓风中心的预测和气象预报不断地取得进步，毫无疑问，这种进步挽救了许多人的生命。

然而，并不是每个人都收听气象预报。约有 80 000 名（几乎占据当时该市总人口的 1/5）新奥尔良人没有来得及撤离，其中 1 600 人失去了生命。对幸存者的调查显示，2/3 的人没想到这次风暴会那么严重，其他人则说当时已经被混乱的撤离秩序搞晕了。尽管梅菲尔德和其他政府官员苦苦哀求，市长雷·纳金还是等了 24 个小时才下令强制撤离。还有一些居民，如穷人、老人或是看不到新闻的人，即使曾经想过撤离，最终也没能逃过这一劫。

本书中成功预测的故事之一就是气象预报，人与机器通力合作，使得我们能够了解，有时还能预测到大自然错综复杂的奥秘。然而，能够预测自然规律并不意味着我们可以改变它。如果没有人愿意相信预测，那预测也就没有什么意义了。卡特里娜飓风就是有关人类才智与失误的故事。

我们真能准确地预测天气吗？

科罗拉多州博尔德市的美国国家大气研究中心（NCAR）有几个超级计算机实验室，它们会自己制造天气：有时炎热，由国际商业机器公司（IBM）公司蓝火超级计算机每秒进行的 77 万亿次运算会产生大量的辐射能；有时多风，所有的热量都需要被冷却，以免预测天气的能力突然丧失，所以，会有几个高压风扇一直向计算机供氧；有时还很聒噪，风扇的噪声很大，而这些运行的声音正是衡量这些装备是否安全的标准。

蓝火超级计算机由 11 个小匣子组成，每个小匣子大概 2.4 米高、0.6 米宽。在匣子的边缘，有一条亮绿色的赛车条纹。在匣子的后端，和你想象中的一样，大量缆线交错，连接着机器的主干，蓝色的指示灯一闪一闪。从匣子的前面看，它们的形状大小与便携式马桶差不多，只不过多了一个带有银色把手的门。

理查德·罗夫特博士是美国国家大气研究中心技术发展部门的主管，他负责监督超级计算机实验室的运行。我与罗夫特博士聊天时说：“这些计算机看着有点儿像马桶。”

气象领域的人已经习惯了这些玩笑。出演《抑制热情》的拉里·戴维就指出，气象专家有时在没有雨的天气也会预测有雨，好像只有这样做才可以在高尔夫球场上领先别人一样。政治广告常以天气打比喻，象征对手在某些问题上总是出尔反尔，以此攻击对手。大部分人认为，气象预报员其实并不怎么擅长预测工作。

事实上，看看那些正在运行中的计算机，想想所有这些操作是不是压根儿没用：计算机就能预测天气？还是它们仍然无法告诉我们明天会不会下雨？

罗夫特看上去并不满意预测现状。改良的计算机运算能力并没有明显提高对地震和经济形势的预测准确率。但是，气象学领域已经取得了许多显著的进步，罗夫特的超级计算机就是功臣之一。

气象预报简史

回到办公室之后，罗夫特对我说："请允许我偏离正常的飞行计划。"罗夫特很幽默，想法古怪离奇，很像电视剧《办公室》里面的德怀特·斯格鲁特，只不过他更有自知之明。罗夫特说，有史以来，人们一直在试图预测自己周边的环境。"早在查科峡谷或是巨石阵形成的那个时期，人们就已经认识到自己可以预测一年当中时间最长和最短的一天，可以预测月亮的运动轨迹。但仍有一些事情古人无法预知，如来自某种动物的伏击或洪水和雷阵雨的突然袭击。"

现在，提前几天预测到飓风是很平常的事，但是气象学发展成为科学的过程并非一帆风顺。几个世纪以来，气象学发展都处于停滞状态。被称为"天文学家"的巴比伦人做出的气象预报，被刻在石碑上保留了 6 000 多年。然而，最终他们还是归顺了宁吉尔苏。亚里士多德写过一篇关于气象学的论文，提出了一些可靠的直觉感知，但不过是浅尝辄止。而仅在过去的 50 年前后，随着计算机的发展，气象学才取得了一些真正的进步。

你可能认为气象预报运用不到玄学知识，但是预测天气这一思想却引起了关于预定论和自由意愿的长期争论。罗夫特问我，"是万物天注定呢，还是万物人注定呢？这是人类最基本的问题，而且这两种思潮确实存在。"

罗夫特继续说道，预定论来自圣·奥古斯丁和加尔文主义，基于他们的哲学理论，人们也许有能力预知未来，但却无法改变未来，所有事都要顺应上帝的旨意。"这与耶稣和托马斯·阿奎奈信仰的自由意愿相悖。这是一个有关世界可预知与不可预知的问题。"

在启蒙时代和工业革命时期，关于不同事物可预知性的争论就已经

开始了。艾萨克·牛顿的经典力学理论似乎证明了宇宙的高度有序性和可预知性，并且遵守相对简单的物理定律。几个世纪以前，科学进步、技术进步和经济进步就像是天方夜谭，而随着“人定胜天”思想的兴起，这些进步意识开始出现了。预定论也成了一个新思想的一部分，这个新思想就是科学决定论。

科学决定论表现为多种形式，但仍无法与法国天文学家、数学家拉普拉斯的观点相提并论。1814 年，拉普拉斯做出以下假设，后来这些假设被称作拉普斯的恶魔：

> 我们可以把宇宙的现状看作其过去已经发生的事情和未来可能发生的事情共同影响的结果。假设我们具备一种理解能力，能在某一个特定时刻认识到使大自然运动的所有力量，能够知晓构成大自然的所有事物的位置。若这种理解能力足够强大，可以对所有这些数据进行分析，就必然能够用一种最简单的公式或准则涵盖这个宇宙中最大的星体和最微小原子的所有运动。有了这种理解能力，就没有什么是不能确定的，未来和过去都能尽收眼底。

若对现在的情况（构成大自然的所有事物的位置）有充分的了解，对统治宇宙的定律（使大自然运动的所有力量）了如指掌，我们就应该能够做出完美的预测（未来就会像过去一样呈现在我们眼前）。宇宙中所有粒子的运动跟台球桌上那些球的运动一样，是可以预测的。拉普拉斯坦言，人类可能还不具备这样的预测能力。但是，如果人类足够聪明（或是有一台高速计算机），还是可以对天气或其他所有事物进行预测的，这样我们就会发现，大自然本身是多么完美。

拉普拉斯的恶魔存在的 200 年里，对它一直争议不断。与决定论者争论的是或然论者。或然论者认为，宇宙可知性成立的条件是，承认一定程度的不确定性的存在。最初的或然论几乎都是认识论范式，认识论断言，人类认识宇宙的能力有限。近期，随着量子力学的发现，科学家和哲学家开始怀疑宇宙自身的运行是否也存在概率。当你仔细观察时会发现，拉普拉斯试图识别的粒子呈波状运动，似乎没有固定的位置。如果一开始就无从得知某物在哪里，你又何以预测它将去向何处呢？显然做不到。这是理论物理学家沃纳·海森堡提出的著名的“测不准原理”（又称“不确定性原理”）的基础。物理学家以各种方式来解释“测不准原理”，但这一原理表明，照字面意思来看，拉普拉斯的假设就不可能正确。如果宇宙本身是杂乱无章的，那么根本就不可能存在完美的预测。

幸亏在研究天气方面用不到量子力学，而只涉及分子（而不是原子）层面，分子体积相对较大，不会受到量子的影响。此外，很长一段时间以来，我们对基于化学和牛顿物理学的天气研究一直都非常熟悉。

那么，将拉普拉斯的恶魔修正一下会怎样呢？如果我们知道地球大气层内每一个分子的位置——比起了解宇宙中每一个原子位置的要求，这个要求低得多——我们能否做出完美的气象预报呢？或者说，天气本身也存在一定程度的不可测性吗？

用矩阵来预测天气

对天气状况进行单纯的统计预测，早已成为可能。假设今天下雨，那么明天也下雨的可能性有多大呢？气象学家可以查看数据库里有关过去下雨的实例，或者查看较长时期内下雨的概率的平均值，如 3 月份的伦敦基本有十一二天都在下雨，进而得出答案。

问题是，这类预测的用处并不是很大，因为它们不够精确，无法告诉你是否需要带伞，更不用说预测飓风的路径了。所以，气象学家一直在寻找其他形式的预测。气象学家需要的是栩栩如生的预测模型，能够逼真地模仿自然界天气变化的过程，而这些是统计预测做不到的。

然而长期以来，我们预测天气的能力远远落后于对天气的理论研究。我们知道如何解方程式，并且能得出正确的结果，但是我们却无法使用方程式计算大气层里的每一个分子，我们能够做到的，只是给出近似值。

将问题简单化往往是最直接的方法，把大气层分解为一系列有限的象素，气象学家通常把这些象素称为矩阵：菱形格或方形格。罗夫特说，成果丰硕的英国物理学家刘易斯·弗莱·理查德森在 1916 年第一个做出这样的尝试。理查德森试图预测某一个特定时间——1910 年 5 月 20 日下午 1 点——德国北部的天气状况。严格地说，这算不上是预测，因为理查德森选择的是 6 年前的时点，但是，理查德森准备了很多数据：由政府收集的关于温度、气压和风速的一系列观察报告。他还有大量的时间：当时，他正在法国北部当救护志愿者，远离战火纷飞的前线。他把德国分成若干个二维分区，每个分区跨越 3 个纬度（约 338 公里）和 3 个经度。之后，他试着解出控制每一分区天气的化学方程式，并算出它们会对相邻地区的天气产生什么样的影响。

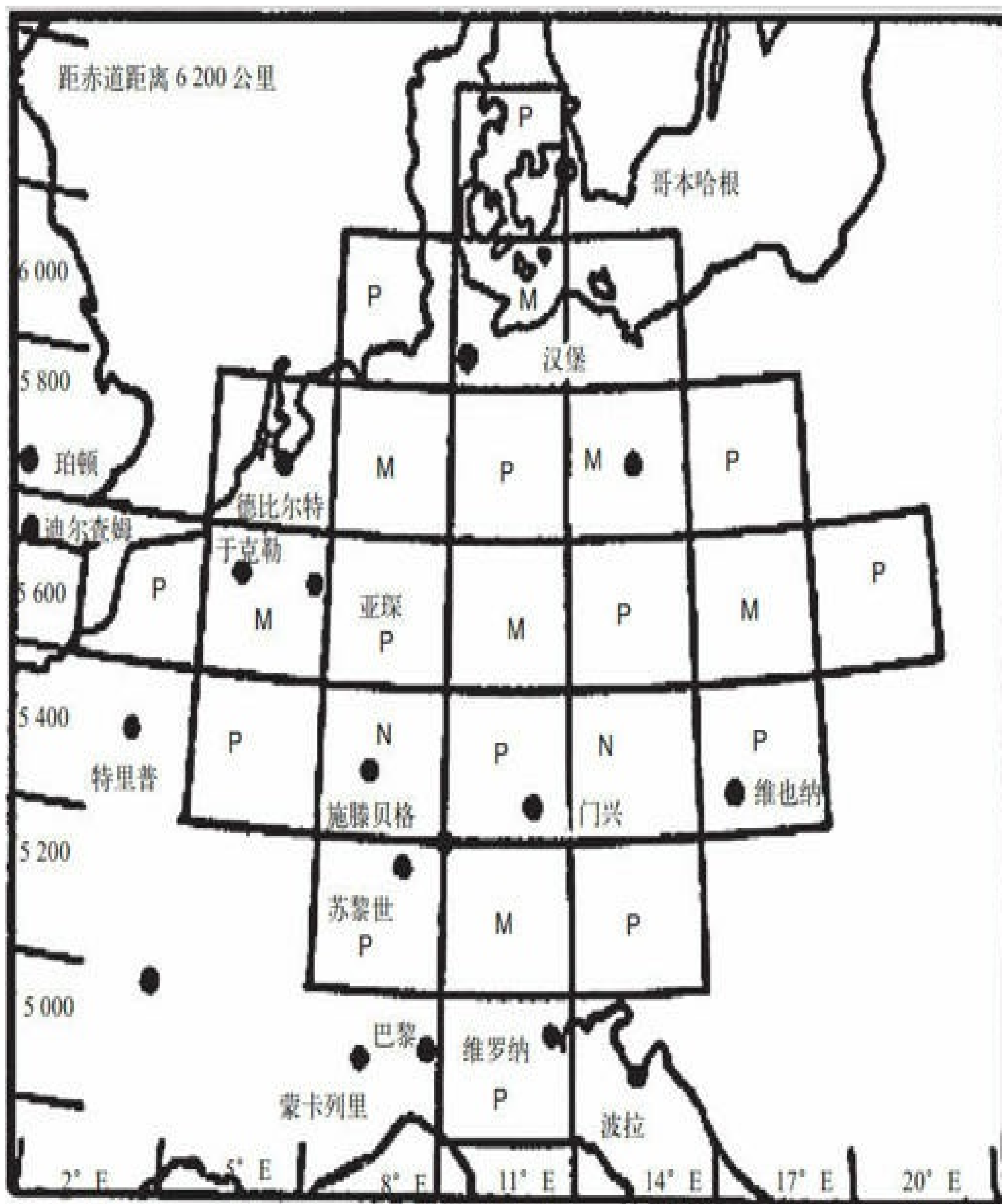


图 **4-1** 理查德森矩阵：现代气象预报的诞生

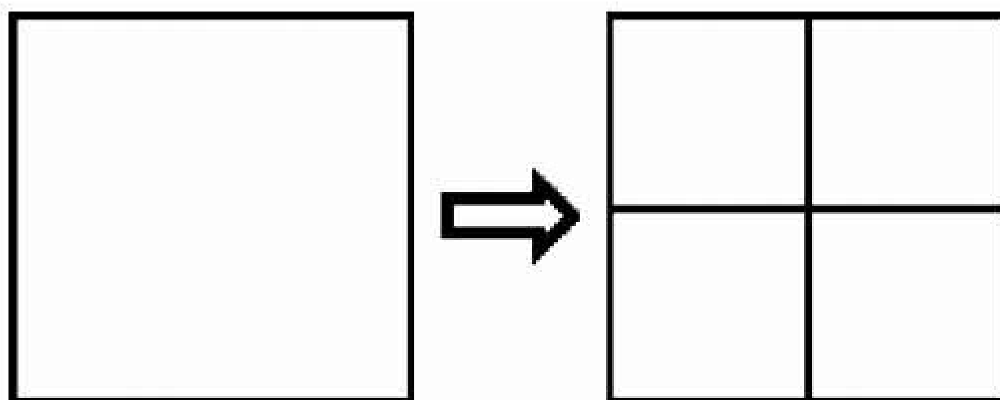
可惜，理查德森的实验并未成功，他预测当天气压会急剧升高，可实际上并没有。尽管如此，理查德森还是公布了这次实验的结果。这似乎是预测天气的正确方法：不依靠粗略统计出来的近似值，而是从第一手资料入手，利用对系统运行的透彻的理论认识进行预测。

理查德森采取的预测方法的问题在于，他需要太多的准备工作。计算机更适合他所建立的这一模型的要求。

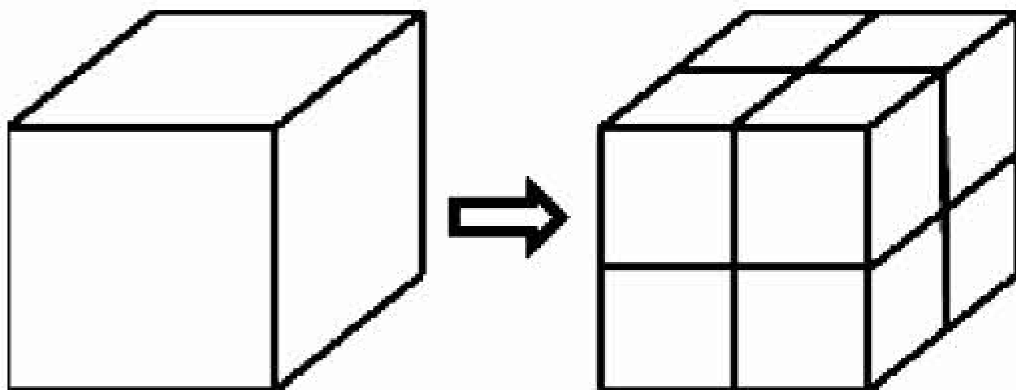
首次使用计算机预测天气是在 1950 年，数学家约翰·冯·诺依曼使用一台每秒可以进行 5 000 次运算的机器，速度远远快于在法国干草堆里用笔和纸做计算的理查德森。但是，这次预测的结果并不好，还不如随意猜测得出的结果。

到了 20 世纪 60 年代中期，计算机才开始展示出预测天气的技能。蓝火是比第一台计算机的计算速度快 150 亿倍，比理查德森几乎快 1 000 万亿倍的超级计算机，超快的计算速度使蓝火看上去更为敏锐。现在的气象预报比 15~20 年前的准确多了，但是，相对于不断提高的运算能力，气象预报准确性的提升就显得有些缓慢了。

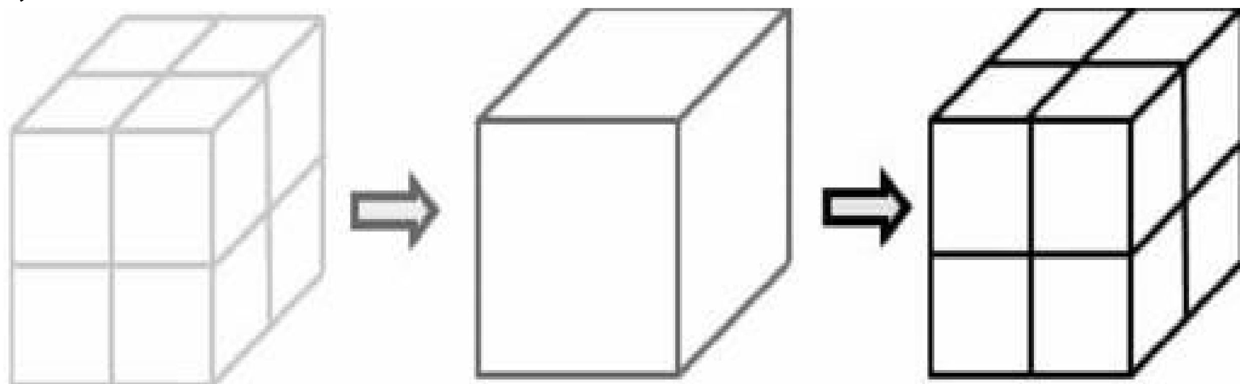
之所以进步缓慢，主要有两个原因。其中之一是，这个世界不是一维或二维的。提高气象预报的准确性，最可靠的方法就是，减小用来代指大气层的网格区域，逐一分析每一个分子的运动。理查德森的单位研究区域大约是 40 000（200 英里×200 英里）平方英里（约合 10 360 平方公里），只能提供一个高度概括的视界（你几乎可以把天气状况完全不同的纽约和波士顿都塞进这个 40 000 平方英里的区域里）。如果把这个单位研究区域的边长减半，也就是让分辨率变为 100 英里×100 英里，那么预测的精确度就会提高，但同时方程式也会增多。实际上，方程式的数量不是增加一倍，而是增加 3 倍。这就意味着，你需要大约 4 倍的计算能力，才能得到答案。如下图所示。



不只需要考虑二维空间，二维以上的空间也需要投入更多精力。不同的气候模型可以停留在较高或较低的大气层中，也可以停留在海洋里或地球表面。在三维空间里，如果单位研究区域的数量增至两倍，方程式就需要增至 8 倍。如下图所示。



接着还有第四维：时间。气象模型如果是静态的，可不见得就是好事，因为我们想要了解的恰恰是天气的动态变化过程。暴风雨的移动速度约为每小时 64 公里：如果是在一个 40×40×40 的三维网格中，就需要每小时观测一次，进而监控暴风雨的移动；但是，如果三维网格是 20×20×20，那么暴风雨每半个小时就会穿过其中的一个区域。这就意味着，如果时间参数减半，计算量就会增加到原先的 16 倍。如下图所示。



如果这是唯一的问题，它不会成为气象预报准确性提升的障碍。一般来说，若想要让单位研究区域的分辨率翻倍，那么你需要掌握 16 倍于原先的运算能力，而运算能力会以每两年增加一倍的速度不断提高。也就是说，只需等待 8 年，得到的预测就会比原先准确一倍。顺便提一下，美国国家大气研究中心的超级计算机已经达到这个运算速度了。

如果你解决了控制天气系统运动的流体动力学法则问题，会怎样呢？相对来说，这些属于牛顿学说的范畴。“测不准原理”对物理学家而言乐趣无穷，对你而言也算不上什么烦恼。你拥有蓝火这样最先进的设备，雇用了理查德·罗夫特为你设计计算机软件并进行模拟实验，还会有什么问题吗？

混沌理论与蝴蝶效应

出问题的是什么呢？是混沌理论。你可能听说过蝴蝶效应：巴西的

蝴蝶扇扇翅膀，就掀起了得克萨斯州的龙卷风。混沌理论是麻省理工学院的爱德华·洛伦兹在 1972 年发表的一篇论文的题目，当时，洛伦兹刚刚开始做气象师的工作。适用混沌理论的系统，常有以下两个特性：

1. 该系统是动态的，这就意味着当前某一个时间点发生的动作会影响未来的动作。

2. 该系统是非线性的，这就意味着其会呈指数型增长而非加法累积。

动态系统给预测者们带来了大量问题，例如，我在本书后文中会提到一件事，美国经济持续发展的连锁反应事件，正是其难以预测的原因之一。非线性系统也是一样，抵押证券引起的金融危机就是这样一个例证，宏观经济中的微小变化极有可能导致严重的后果。

如果把这些特性放在一起，简直就是一团糟。洛伦兹一直没有意识到这些问题的严重性，他的重大发现只是出于偶然，就好像亚历山大·弗莱明意外发现了青霉素、纽约尼克队出现了林书豪一样。

洛伦兹和他的团队早期用一台名叫皇家麦克比 LGP–30 的计算机制作了一套气象预报程序，他们本以为会有所收获，可计算机后来却给出了很多稀奇古怪的结果。他们使用自认为完全相同的数据和完全相同的操作代码，但这套程序给出的堪萨斯州的气象预报却一会儿晴一会儿阴，每测一次出现一个结果。

经过几周反复的硬件检查和程序调试，洛伦兹和他的团队最终发现，他们使用的数据实际上并不完全相同：一位技术员把数据精确到了小数点后三位。比如，网格中某一角上的气压本该是 29.516 8，却被写成了 29.517。这怎么可能不造成巨大的差别呢？

洛伦兹认识到小数点后的数字保留情况会造成巨大的差距。混沌理论最基本的信条是，初始条件的一丁点儿变化，比如巴西的蝴蝶扇动翅膀就会产生巨大的、无法预料的各种结果，会引发得克萨斯州的龙卷风。但这并不是说这一系统就像它的名字“混沌”的含义那样随机，“混沌理论”也不是“墨菲定律”（该出的错总要出）的现代版本，它只是意味着，某些类型的系统很难预测。

我们的数据中一旦出现错误（或者假设中出现错误，如抵押贷款证券的例子），问题就会随之而来。想象一下，我们本该计算 5 加 5，可是键入第二个数字时出错了，变成了 5 加 6，我们本来想得出的结果是 10，现在结果却成了 11，这必然会出错，但错得还不算离谱：加法是

线性运算，不会出现让人无法原谅的错误。而如果指数运算出现错误，后果就会非常严重。如果本该计算 55，却误写成 56，那得到的结果就从原来的 3 215 变成了 15 625，这样就错得太离谱了，几乎是正确答案的 5 倍。

如果这个计算过程是动态的，就意味着我们在整个过程中的某一个阶段的输出会成为下一个阶段的输入，结果的不准确性就会更加严重。比如说我们要算 5 的 5 次方，然后对得到的结果再进行 5 次方运算。如果犯了上述错误，把第一个 5 次方错写成 6 次方，又把错误结果应用到后面的计算，这样一来，得出的结果就会是应得结果的 3 000 多倍。一个小小的，甚至看似微小的错误，就这样越错越离谱了。

天气变化是动态系统的一个缩影，反映大气层里气体和液体运动的不同方程式是非线性的。因此，混沌理论绝对适用于气象预报，但也使气象预报更容易受信息不准确性的影响。

有时，错误的产生是人为的结果。更为重要的问题是，在我们观察周围事物时，精确程度是有限的。没有哪个温度计是准确无误的，就连小数点后保留位数的不确定都会对预测造成严重的影响。

图 4–2 展示的是欧洲天气模型的 50 轮预测结果，该模型试图预测法国和德国在 1999 年平安夜的天气状况。所有这些模拟操作使用的都是同一款软件，做出的天气假设也是相同的。实际上，这些模型完全是决定论的：他们认为，只要非常清楚初始条件，就一定能非常准确地预测出天气状况。但是，输入信息的一丁点儿变化就会造成输出信息的巨大差异。欧洲天气预测模型曾经试图解释这些错误。在模拟操作中，汉诺威市的气压也许只作了微调，斯图加特市的风力也许只变更了零点几，这些微小的变化都足以引发巴黎的强风暴，而其他地区可能仍是平静的冬夜。

这就是现代气象预报的制作过程，人们有意加入这些微小的变化以代表观测数据内在的不确定性，这样就可以将一个必然性的预报变成了或然性的。比如，当地气象预报员说，明天有 40%的概率会下雨，意思是说，在他的模拟操作中，有 40%的概率会下雨，有 60%的概率不会下雨。

然而，问题并没有那么简单。气象学家使用的气象预报程序相当不错，但也并非尽善尽美。我们实际看到的预测反映的是计算机和人工判断相结合的结果。人类可以使计算机预测得更好，也可以使计算机预测得更糟。

视觉化预测与抽象化预测

世界天气和气候预测大楼是典型的 20 世纪 70 年代建造的办公楼，造型很一般，外观是黄褐色的，位于马里兰州的坎普泉，距离华盛顿有 20 分钟路程。这座大楼是美国国家海洋和大气管理局（NOAA）的指挥总部，在政府组织关系表中，NOAA 是国家气象局（NWS）的上级单位。相比之下，位于博尔德的国家大气研究中心（NCAR）则可以使人们饱览落基山弗兰特岭地区的景观，而世界天气和气候预测大楼给人的印象不过是一个官僚机构。

美国国家气象局最初是在 1870 年由格兰特总统授权，由战备部组建的。之所以由战备部组建，一部分原因是因为格兰特总统相信，只有保持军事纪律的作风，才能保证预测的精准，而另一部分原因是，当时的整个战备部都死气沉沉的，只有在竭力取胜的战争时期才有存在的价值。



1888 年 1 月发生“校舍暴风雪”事件之后，公众开始更多地关注气象

预报问题。那年 1 月 12 日，起初还算是大平原地区相对温暖的气温，可几个小时之后，气温骤降 30 摄氏度，紧接着，让人眼晕的暴风雪骤起。数以百计的学生刚刚放学就被暴风雪困住，冻死在回家的路上。早期的气象预报如此粗糙，但对于这种严酷的天气情况，人们还是期望气象预报至少能提供一些预警。于是，美国国家气象局被划归农业部管辖，接手一些面向大众的任务。

现在的美国国家气象局仍保留着最初的职能，预测者夜以继日地工作，薪水却不高，因此预测者自称“人民公仆”。我在坎普泉见到的气象学家都是爱国主义者，他们总会让我由衷地认为，农业、小商业、航空公司、能源部门、军事、公共服务、高尔夫球场、郊游野餐以至上学的小学生之所以能够一直保持这样生龙活虎的运行状态，气象预报起到了重要作用，预测者更是劳苦功高（尽管美国国家气象局对国民经济会产生 20%的直接影响，但每年得到的经费却只有 9 亿美元，相当于每个美国市民付出 3 美元）。

我见过的气象学家中，有一位名叫吉姆·呼克，是美国国家气象局水文气象预报中心的主管。他在这一领域工作了 35 年，既负责计算方面的工作（协助建立预测者使用的计算机模型），也负责操作方面的工作（做出预测并把预测结果公之于众）。对于人类和机器如何探索气象领域，他有着独特的见解。

与能够进行每秒 77 兆次浮点运算的计算机相比，人类究竟哪点做得好呢？答案是，人类能够看得见。呼克让我站上预测台，预测台是由一系列工作站组成的，工作站上悬挂着许多蓝色标牌，上面配有海事预报中心和国家中心等机构的传奇故事。每个工作站由一两名气象员进行操作，配有一套平板屏幕显示器，显示器上的彩色地图展示了美国每个角落的各种天气数据类型。预测者工作的时候，悄然无声却又动作迅捷，有一种类似格兰特军队的精确作战风范。

有些预测者用光笔在这些地图上做标记，仔细地校准计算机模型给出的温度梯度的轮廓，向西跨越密西西比三角洲约 24 公里，向北覆盖伊利运河约 48 公里，离他们期望中的那份近乎完美的气象图更近了一步。

预测者知道计算机模型的纰漏。这些纰漏不可避免，因为这是混沌理论影响的结果。即使是模型中最微小的漏洞，也可能造成潜在的深远影响。比如，当普吉特海湾出现低压系统时，计算机对西雅图夜间大雨的预测也许会过于保守；当缅因州的阿卡迪亚国家公园出现大雾时，如果大风朝着某一个方向吹，黎明时大雾就会散去，可是如果风来自另外一个方向，那么雾需要一个上午才能消散，计算机也许预测不到这一

点。这些经验都是预测者在试图对付计算机模型中的纰漏的过程中逐渐积累起来的，这就好比技艺高超的台球选手能把台球桌上死角位置的球，调整到顺应自己球杆的位置。

这些气象预报员运用的独特资源，就是他们的视觉。对所有学科的预测人员而言，视觉都是十分有用的工具。对显示两个变量互相作用的图表进行目视检查，通常能比统计测试更快捷、可靠地检测出数据中的异常值。这也是计算机远远落后于人脑的地方之一。若对一串字母稍加变形——就像用于垃圾邮件拦截或是密码保护的验证码技术一样——“聪明透顶”的计算机也会被迷惑。计算机缺乏想象力，哪怕对操作稍作变动，它就无法识别图案，而人类经过必要的进化，拥有十分强大的视觉皮质，能够快速识别任何数据变化，辨认出抽象的图案或组织。在不同类型的天气系统中，抽象的特质恰恰非常重要。



图 **4-3** 验证码

其实，在气象计算机作用不大的过去，气象预报几乎完全依靠视觉判断。那时候没有平板屏幕，气象办公室里有的只是一排排的看板台和照明地图，气象员用白垩笔和绘图铅笔在上面做记号，一次只能预测出方圆 24 公里的天气情况。尽管看板台在许多年前就被淘汰了，可这项技术所需要的一丝不苟的精神却一直延续至今。

呼克解释道，最好的预测者在进行视觉化和抽象化思考的同时，还要能够整理计算机提供的大量信息。另外，对于所研究的系统，他们还必须理解其动态和非线性的属性，这并非易事，需要充分调动左右脑的功能。呼克手下的许多预测者本来都能加入高级工程师或软件设计师这类高收入群体，可他们却选择了从事气象工作。

美国国家气象局保存着两类图书：一类展示了计算机的风采，另一类展示了人类的贡献。根据该机构的数据来看，在单独由计算机指导完成对降雨和温度的预测后，人类还能将其精确度分别提高约 25%和10%。另外，据呼克说，多年来这些数据一直保持着平稳态势：尽管计

算机技术取得了长足进步，但他手下的那些预测者仍不断为其添彩，视觉判断功不可没。

被雷电击中的概率越来越小了

20 世纪 70 年代中期，呼克开始了从事气象预报工作的职业生涯，那时关于气象预报的玩笑话其实还是有些道理的。比如，一般来说，国家气象局在 3 天前做出的气象预报，难免会将高温报高或报低 6 摄氏度（见图 4–4）。如果我们自己去查长期的天气变化表，得出的结论也差不多是这样。好在人机合作带来了很多好处，现在的平均误差是 3.5 摄氏度，降低了将近一半。

气象预报员对恶劣天气的预测也越来越准确了。你知道遭雷击致死的概率是多少吗？实际上，这并不是一个常数，是否在雷雨天气出门，或是在气象预报不准时能否及时避雨，这些都会影响遭雷击的概率。1940 年，美国人遭雷击死亡的概率是 1/400 000，今天，这一概率只有 1/11 000 000，概率下降了近 3000%。这一变化一方面反映出我们生活方式的改变（我们的室内工作越来越多）以及通信技术、医疗保健水平的改善，另一方面也得益于越来越准确的气象预报。

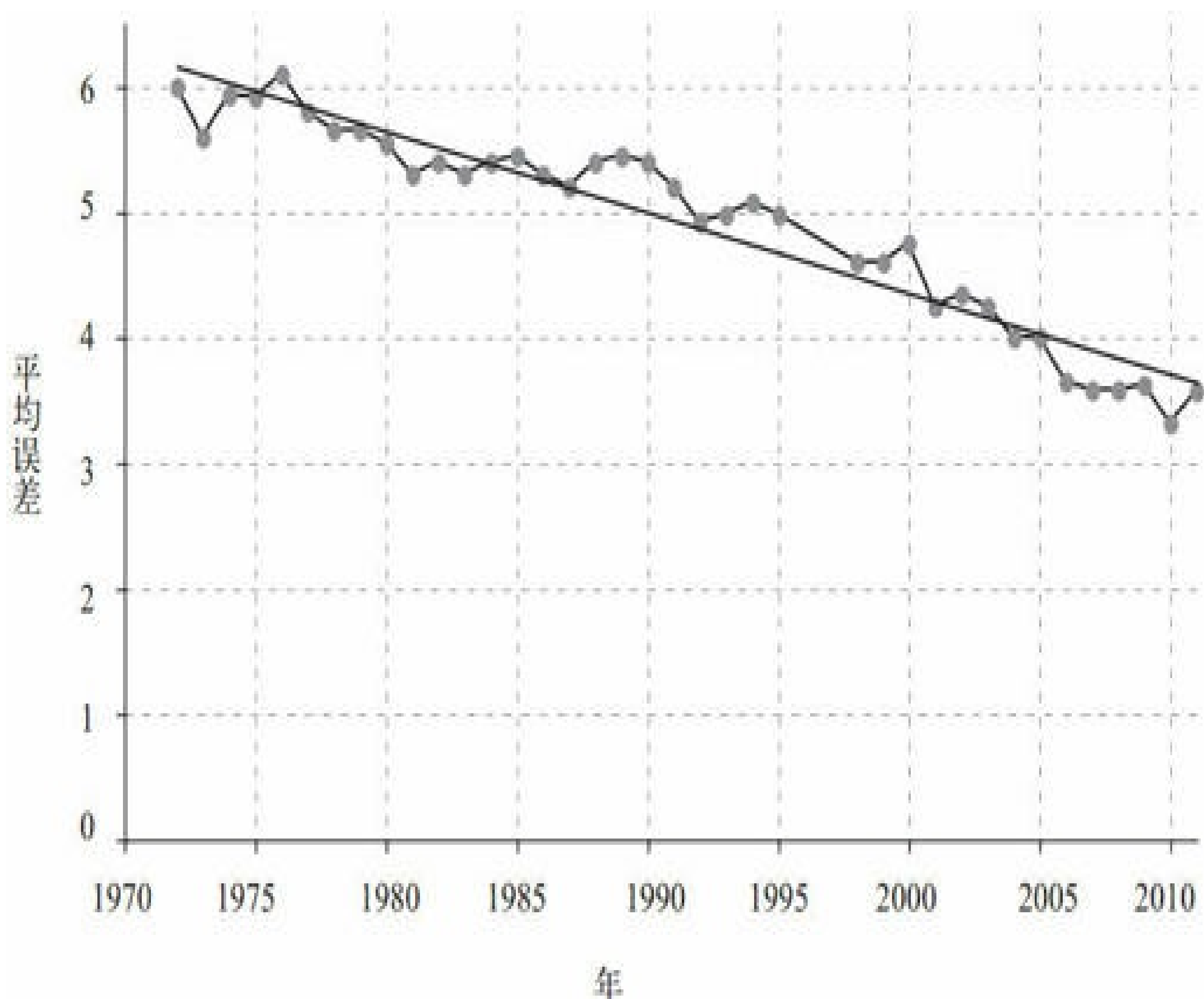


图 **4–4** 国家气象局的高温误差平均值

给人印象最深刻的进步大概就是飓风预测了。25 年前，美国国家飓风中心提前 3 天预测飓风登陆的位置时，平均误差将近 563 公里，对于人类来说，这种预测毫无意义。比如，以新奥尔良为圆心以 563 公里为半径向外扩散，涉及的地区包括休斯敦、得克萨斯、塔拉哈西和佛罗里达（见图 4–5），而我们根本无法疏散这么大区域的居民。



图 **4–5** 改进后的飓风轨迹预测图

现在的平均误差只有约 161 公里，覆盖的区域只有路易斯安那东南部和密西西比南部的一小片地区。虽然飓风仍会在某一个时间到达这一区域以外的地区，但我们现在关注的区域相对较小，预测的效果也能相对好些，足以提前 72 个小时疏散那里的居民。1985 年，对飓风登陆时间的预测只能提前 24 个小时，与之相比，现在在风暴来袭之前，我们有近 48 个小时的预警时间，对于新奥尔良这样的城市，疏散工作到后期，时间就越显得宝贵。[1]

美国国家气象局没有否定拉普拉斯的恶魔，你可能会认为他们应该

得到的赞誉比实际得到的要多。尽管复杂多变的天气系统对预测构成了巨大挑战，但气象预报科学仍然大有斩获。读完本书你会发现，在进行预测时，这样的成功案例更多的是例外，而非规律。（还是把你的玩笑话留给那些经济学家吧。）

事实上，美国国家气象局的工作通常得不到人们的认可，面临着来自私有行业的严峻竞争，这种竞争是在不公平的运动场中进行的。美国国家气象局的模型数据免费提供给所有需要它的人，而不是采取世界上其他地区的大多数相应机构的做法（其他大部分国家的气象局会收取预测许可费和使用费）。这样一来，像美国 AccuWeather 气象预测公司和气象频道（TWC）这样的私有公司才能够顺带发展并销售自己的产品。大量的受众会从这些私有公司中的某一家获取预测信息，气象频道网站 Weather.com 的浏览量比政府气象预测网站 Weather.gov 的浏览量多出近 10 倍。

我通常很赞同自由市场的竞争，或是国有企业和私有企业之间的竞争。棒球运动迅猛发展，得益于棒球预测更好地结合了球探和数据怪才在预测潜力球员发展前景时的见解，这都是竞争带来的结果。

然而，在棒球运动预测领域中，衡量竞争力的标准十分明确：你赢了几场比赛？在气象预报过程中，问题则有些复杂，国有和私有预测部门各自会有不同的衡量标准。

什么样的预测才算是好预测？

美国气象频道最重要的科学家兼副总裁布鲁斯·罗斯博士为人谦和，他告诉我：“没有人碰到过专业的研究专家看气象频道，但其实他们中有很多人都是偷偷看的。”罗斯并不认同气象频道的预测比政府部门发布的预测更精确，他只是说，二者的预测存在差异，是为各自的特定受众群提供的。

罗斯还说道：“通常人们不会根据这些模型对实际天气的预测质量衡量气象预测结果。真正重要的是，纽约市的天气是 25 毫米的降雨，而不是 254 毫米的降雪。因为对于普通受众来说，这有巨大的差别，而科学家们对此却不以为然。”

的确，罗斯大部分的时间都致力于研究受众如何理解自己的预测这类问题，这些问题高度实用，甚至还有些陈腐。比如，如何将第一手的天气数据转化为日常用语：严寒到底有多冷？当天下暴雨的概率有多大？少云和多云之间的界限该如何划定？这是气象频道需要解决的问题，因为该频道发布过太多预测，决意在措辞上有些变通，所以现在应该建立一些解决问题的正式规则。

有时，让预测适应受众的需要，这种做法显得滑稽可笑。多年来，气象频道一直用绿色阴影在雷达地图上标注降雨（偶尔也会用表示大暴雨的黄色和红色标记）。2001 年，销售部的一个员工想到一个新点子：用蓝色代替绿色标注降雨，毕竟这才是人们眼中水的颜色。结果，气象频道马上被气愤的——偶尔也有惊慌的——受众来电围攻，人们误以为蓝色的标注代表的是至今未知的降雨（血浆暴雨或是放射性尘埃）。罗斯博士对我说："人们的反应就像核爆炸！有人还写信来说道，'这么多年你一直告诉我们雨的标识是绿色的，现在怎么又变成蓝色的了？你们疯了吗？'"

但是，气象频道在气象学方面的态度也很认真。至少在理论上，我们有理由相信它们能够做出优于政府部门的预测，毕竟气象频道能够以政府部门所有的原始数据作为起点，并把它们能够贡献的价值毫无保留地投入进去。

可问题是，什么才是"更好的"预测呢？我简单地将其定义为更准确的预测，但也有一些不同的声音，这些观点也都十分中肯。

1993 年，俄勒冈州立大学的气象学家艾伦·墨菲发表了一篇颇具影响力的文章，他在文中提出假设，在气象预报群体中，有 3 种常见的关于预测质量的定义。墨菲并没有特意指出哪一种定义更好，只是试图促使众人对这几种定义进行更为开放、更为坦诚的交流。这些定义的不同版本几乎可以在任何可预测或可预言的领域中应用。

墨菲提出，衡量预测的一种途径——或许也是最显而易见的途径——就是通过他所说的"质量"，但何种质量才可被认定为准确呢？"质量"是指预测与实际天气相符吗？

第二种途径就是墨菲所称的"一致性"，但我以为应该是"诚实性"。不论预测得多么准确，这个预测是预测者当时的巅峰之作吗？这个预测是否反映了预测者的最佳判断呢？公之于众之前，这个预测是否作了某种程度的修饰呢？

墨菲最后写道，预测是有经济价值的，判断的依据就是，预测是否有助于公众或政治决策者做出更好的决定。

墨菲对准确和诚实所做的区分很细微，也相当重要。当我做出错误的预测时，我总会问自己，鉴于我所了解到的情况，这是不是我本该做出的最好预测呢？有时，我给出的答案是肯定的，因为我的思考过程是合理的，我已经进行了研究，建立了良好的模型，仔细考量了其中的不确定性。当然，有时我也会发现所进行的预测里有自己不喜欢的地方。或许我过于匆忙地搁置了一个关键证据，或许我高估了问题的可预测性，或许我在某些地方有所偏倚，或许根本就是动机不纯。

我并不是建议你每次预测错误时都要狠狠地责备自己，相反的，当你发现事情超出自己的瞬时掌控时，依旧能保持平和的心态，这才是做出正确预测的前提。但仍要留出空间问问自己，当时作决定时，自己的大脑里到底在想什么。

长远来看，当我们动机正确时，墨菲提出的准确性和诚实性就会汇于一点。但有时我们却做不到动机正确，比如，“麦克劳夫伦讨论小组”就肯定通不过墨菲的“诚实性预测”测试，他们似乎更关注自己在电视上表现得是否聪明，而不关注是否做出了准确的预测。他们本可以表现得很理性，但如果想要引起带有政党倾向的观众的注意，或是想要再次获邀参加节目，他们就会蓄意做出错误的预测。

墨菲的第三个途径是，预测的经济价值会使问题更加复杂化。比如，有人很同情罗斯博士的处境。如果预测一个城市的气温接近冰点，其降水形式可能是雨、冰雹或是雪，这真的值得大家给予更多关注，因为这几种情况都会对早上出行和居民安全造成不同的影响。然而，这更应该是气象频道集中资源、下大力气处理的问题。没有必要怀疑预测的准确性和诚实性。新闻报道力求保证其文章内容的准确性和诚实性，但它们仍然需要决定哪些文章可以放在头版头条。气象频道也要做出类似决定，经济因素是促使其这么做的合理理由。

然而，有时候，对于准确性、诚实性和经济价值的追求也会陷入更加激烈的斗争，而此时商业成功的重要性就要超过准确性了。

商业竞争如何使预测变得更糟糕？

任何气象预测都必须通过两项基本测试，以表明其优势所在：

1.它们必须做到气象学家所说的持续性，甚至做得更好。所谓持续性，是指假定明天（或者第二天）的天气同今天一样。

2.它们还要突破气候学规律。气候学规律是指长期以来特定区域、特定日期历史平均状况的综合。

早在理查德森、洛伦兹和蓝火计算机出现以前，我们的祖先就已经在使用这些方法了，如果我们无法改进这些方法，那么再昂贵的计算机也做不出出色的预测。

我们拥有大量有关过去气象预报结果的数据，至少可以追溯到第二次世界大战时期。比如，我可以登录气象网站 Wunderground.com，查询到 1978 年 1 月 13 日（我的出生日期）早晨 7 点，密歇根首府兰辛的天气状况：气温约 8 摄氏度，小雪，东北风。但是，几乎没有人会劳神

费心地收集过去的气象预报。有人料到那天早上兰辛会下雪吗？你可能会到互联网上查找这样的信息，但却查不到。

2002 年，一位名叫艾瑞克·弗娄尔的企业家改变了这种状况，他是俄亥俄州立大学计算机科学专业的毕业生，当时效力于美国第二大长途电话运营商美国世界通信公司（MCI）。最初只是为了比较政府预测模型和私有公司预测模型哪个更准确，他才开始收集国家气象局、气象频道和 AccuWeather 公司发布的有关气象预报的数据。起初，这个大规模的科学实验的项目多半是为了满足弗娄尔的个人求知欲，可随后迅速发展为有利可图的商机。人们创建了 ForecastWatch.com 网站，在很大程度上按照客户的要求对数据进行重新包装，其客户群的涵盖面很广，从能源商人（对这些人来说，气温发生的细微变化能转变为数万美元）到学者不一而足。

弗娄尔发现，没有一家气象预测机构是明显完胜的赢家。他收集的数据表明，AccuWeather 网站在降雨量预测方面的误差最小，气象频道在温度预测方面略胜一筹，而政府预测在各个方面都稳定出众，总之，它们做得都非常好。

但是，这些模型的预测时间跨度越长，其预测准确性就越低（见图 4–6）。比如，提前 8 天的预测几乎没有任何技术含量，这种预测打破了稳定性，但并没有突破气候学规律。如果是提前 9 天或更多时间进行预测，那么这些专业的预测得出的结果就会比按照气候学规律推测的结果还要糟糕。

罗夫特告诉我，混沌理论是大势所趋，大气的动态内存会自我清除。尽管下面这个类比有些不准确，但对于理解这一原理还是有所帮助的。人们把大气想象成纳斯卡赛车的椭圆车道，围绕轨道行驶的不同车辆就代表不同的天气系统。比赛开始的十几圈，了解赛车的出发顺序能使我们更好地预测出它们经过的顺序。期间可能出现碰撞、急停、引擎故障等我们无法解释的状况，所以我们的预测不会完美无瑕，但总会比随便猜测的准确一些。很快的，速度较快的汽车就会领先速度较慢的车若干圈，不久，赛道上的顺序就完全被打乱了。也许与排位第二的赛车并驾齐驱的，是排位第 16 的赛车（即将落下一圈）和排位第 20 的赛车（已经落下一圈，眼看又要落下第二圈）。比赛最初的状态几乎没有什么参考价值，同样的，一旦大气有了足够的循环时间，天气模式与其最初的状态就不再相似，这些模型也就没有什么用处了。

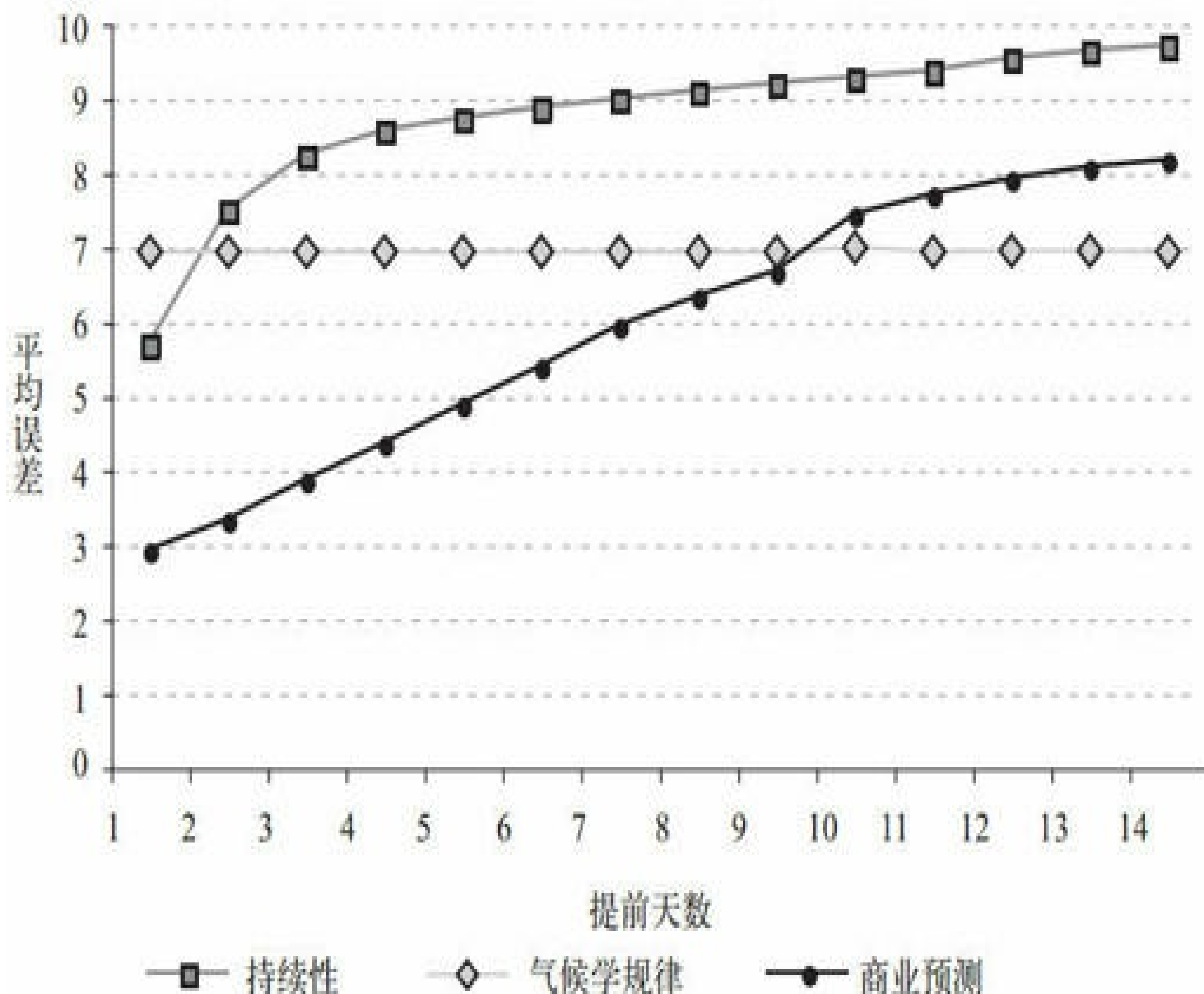


图 **4–6** 高温预测对比

弗娄尔的发现仍然引起了一些令人不安的问题。预报发布的七八天之后，如果计算机模型的预测结果被证明是零技术含量，则会是另外一种情况。这些模型得到的结果竟然还不如普通人坐在家里查阅长期天气平均状况表得到的结果准确，怎么会这样？也许是因为计算机程序对天气系统中自然出现的反馈过于敏感，于是开始自我反馈。这不仅说明噪声中不再有信号，还说明噪声正在渐渐增强。

还有一个更大的问题，那就是如果提前过长时间做出的预测不准确，那气象频道（预测近 10 天的天气情况）和 AccuWeather 网站（将预测时间提前到 15 天）这样的公司为何还要继续发布预测呢？罗斯博士认为，因为这样做不会造成任何伤害，即使是单纯基于气候学的预测，对他们的用户而言也是有点用处的。

对于商业性的气象预报来说，统计学上的准确度没必要斤斤计较。在受众眼中，只有感知上的准确度才是有价值的。

比如，以赢利为目的的气象预报公司很少确切地预测下雨的概率为50%，这个概率对用户来说似乎显得空洞又模糊。相反，它们会投掷硬币且将数值四舍五入，得到60%或者40%的降水概率，尽管这样做会使预测结果更加不准确、不诚实。

弗娄尔还发现预测公司竟然明目张胆地篡改数字，这可能是气象预报行业里公开的秘密了。大多数商业性气象预报都是有偏向性的，可能是故意为之。这些预测公司尤其偏向于预测更多的降水量（但实际上降水并没有那么多），气象学家称之为“降水偏向”。从政府部门得到越多的原始数据，客户面对的气象预报就越多，这种偏向性就会越严重。预测就是这样通过减少准确性来“增加价值”的。

天气预报说降水概率为60%，你出门会带伞吗？

关于预测的重要测试中有一项叫作标定，我认为这是最重要的测试。很多次你都说降水概率为40%，但真正下雨的情况有几次呢？如果长期以来，下雨的概率的确为40%，那就说明你的预测是已标定的。而如果下雨的概率有时只有20%或者高达60%，那么你的预测就是未标定的。

很多领域都难以实现标定，它要求我们做到用概率的方法思考问题，而我们大部分人（包括大部分预测“专家”在内）对此都不是十分擅长。标定会给“过于自信”的预测者——大多数预测者都具有的特点——当头一棒，还需要利用大量数据进行充分评价，也就是对预测者发布的数百个预测进行全面评估。

气象学家的预测涉及大量数据，他们每天都要预测几百座城市的气温、降雨和其他类型降水的概率。一年下来，气象学家得进行数万次预测。

这种频繁的预测不仅在我们想评估某个预测的时候大有帮助，对那些预测者本身也很有用，他们会从中得到很多反馈，知道自己当前的做法是不是有什么不妥，以便根据情况加以改变。比如，某些计算机模型倾向于给出多雨天气的预测——比应有的下雨天气多。一旦察觉到这一偏向，你就可以将其改正过来。同样，如果你对自己的预测过于自信，你很快也会认识到这一点。

美国国家气象局的预测已被证实为极好的标定预测（见图 4–7）。如果美国国家气象局说降水概率为20%，就真的有20%的概率会下雨。美国国家气象局一直充分利用反馈信息，所以预测才真实、准确。

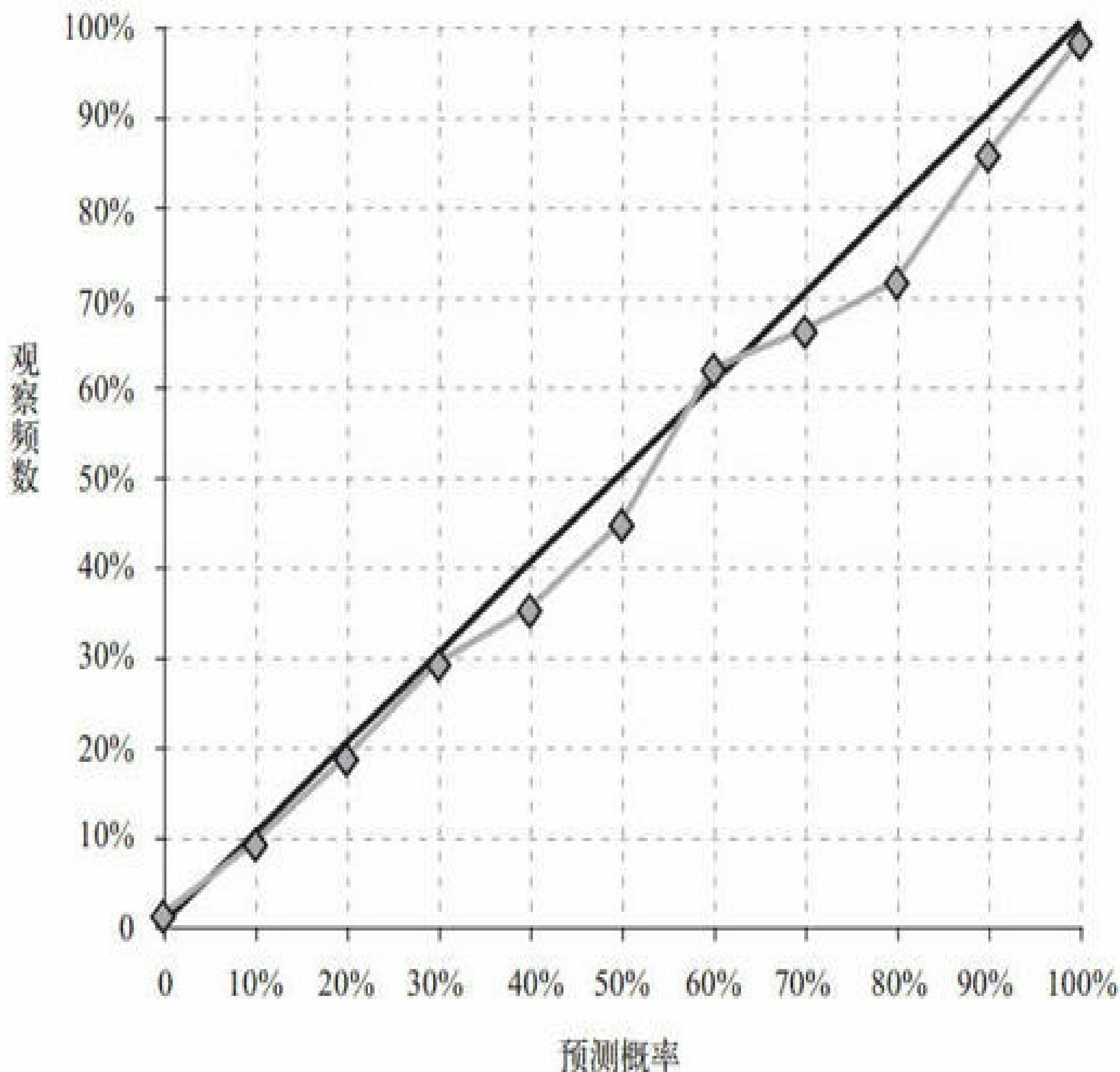


图 **4–7** 国家气象局标定图示

气象频道的气象学家在某些情况下会编造事实。比如，从过去的情况看，当他们预测下雨的概率为 20%时，实际只有 5%的概率会下雨。对此，气象频道承认自己是故意为之，这与气象频道的经济动机有关。

与虚假的天气预报相比，人们对气象预报未能预测到大雨天气的失误更为关注。当大雨不期而至时，人们就会抱怨是气象预报员破坏了他们的野炊，而天气意外放晴则会被视为额外的嘉奖。这并不是科学，但是就像气象频道的罗斯博士对我说的那样："如果预测是客观的，对降水概率的预报没有任何偏向，那我们可能就有麻烦了。"

气象频道仍然是一个相对守旧的组织——它的很多用户误认为它是一家政府机构——大多数时候总是照章办事。他们对降水的虚报仅限于

在不大可能下雨时稍稍夸大降雨概率——当他们知道实际降雨的概率只有 5%~10%时，会假称降雨的概率为 20%——万一露馅了还能为自己遮丑。而有时，气象预报的预测又十分符合标定水准（见图 4–8），比如，当它预测下雨的概率为 70%时，降雨概率还真就是 70%。

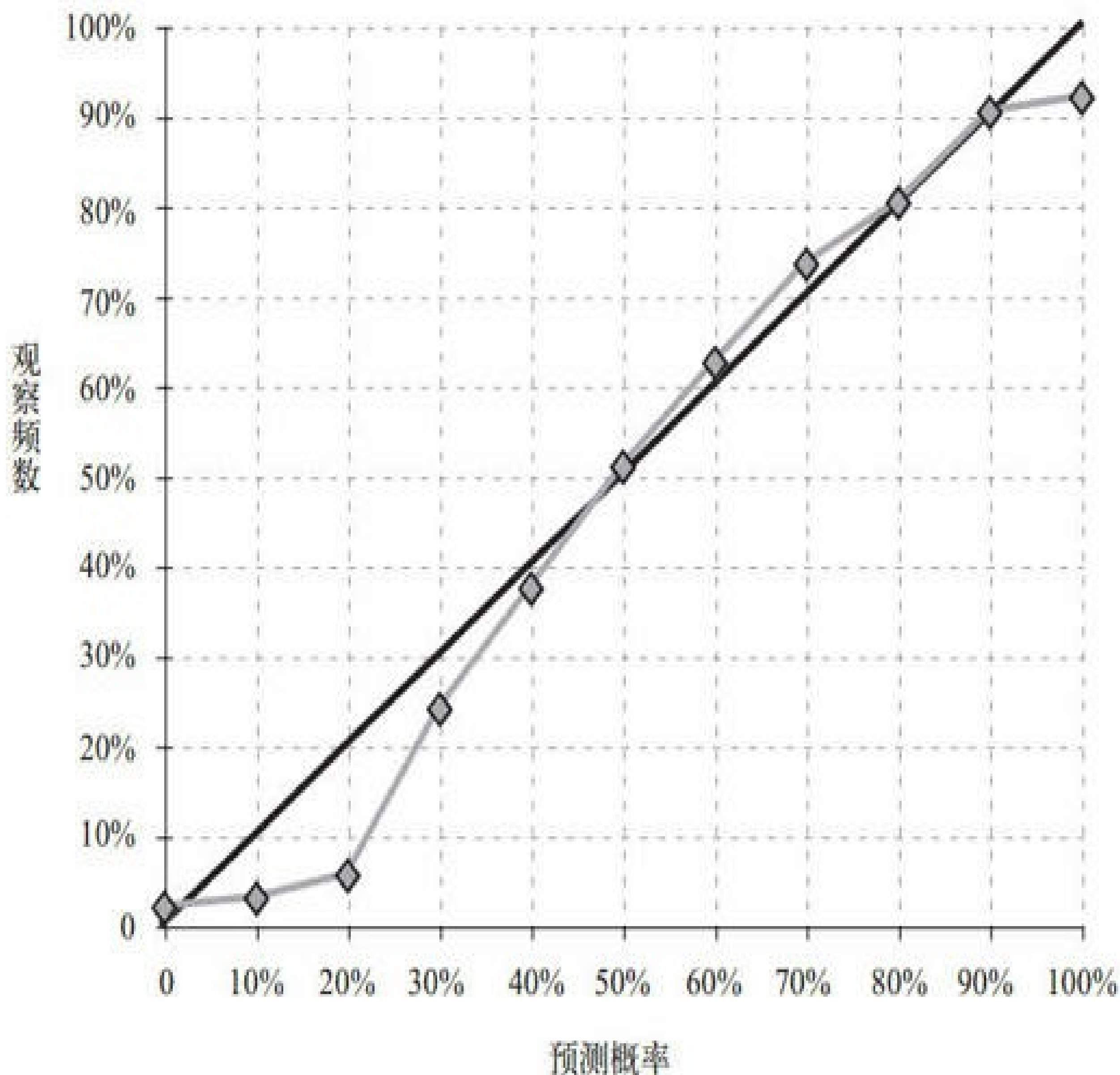


图 **4–8** 气象频道标定图示

若当地的网络新闻中发布了气象预报，事情就会变得混乱不堪。此时，（预测者的）偏向便会突显出来，完全顾不得准确性和真实性了。

堪萨斯城应该算作气象预报的“大市场”了。这里有酷暑寒冬，有龙卷风和干旱，这个城市的气象预报足以登上所有主流网络。一位名叫 J·D·艾格尔斯顿的当地人，为了帮女儿完成五年级的课程项目，开始跟踪分析当地的电视气象预报。艾格尔斯顿发现这项分析非常有意思，所

以连续做了 7 个月，并把结果发送给“魔鬼经济学”博客。

电视气象预报员通常不怎么关注预测的准确性，他们可以利用美国国家气象局通过网络或电台免费发布的预测信息，但是与这些信息相比，他们的预测准确性要差很多，而且远不符合标定水平。在艾格尔斯顿的研究里，当堪萨斯城的电视气象预报员预测某些日子 100%会下雨时，事实上有 1/3 的时候根本没下雨（见图 4–9）。

但是，电视气象预报员不会为此道歉。一位气象预报员告诉艾格尔斯顿：“雇用气象员时并没有对其预测准确性的评估，口头陈述表现反而比准确性更重要。”另一位预报员说：“对于电视观众来说，准确性根本不是大问题。”他们的这种态度似乎都是善意的玩笑：谁会在意这一点点的降水偏向呢？尤其还是为了提高收视率。既然公众并不认为我们的预测有多准确，那我们为什么还要劳心费神地做到准确呢？

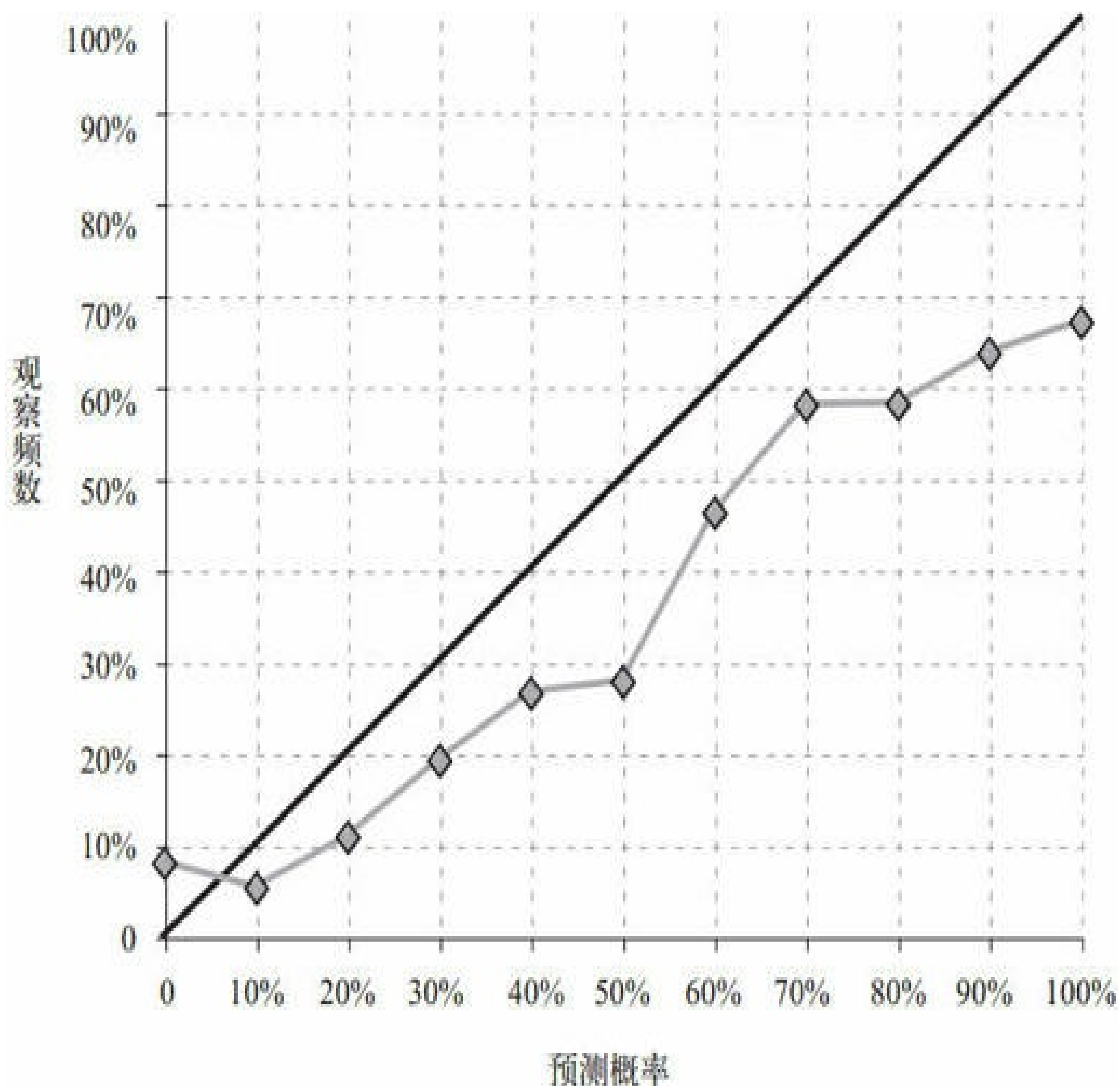


图 **4–9** 当地电视台气象预报员标定图示

这个逻辑有点像“鸡生蛋，蛋生鸡”。电视气象预报员认为他们不必劳神地进行准确的预测，因为他们认为不管怎样都得不到公众的信任；而公众不信任气象预报员，就是因为他们的预报不准确。

当出现卡特里娜飓风这类紧急情况时，彼此信任就成了一个更加严重的问题。很多美国人并不是直接从飓风中心得到的信息，而是从当地的信息渠道获得的，因此，他们还会继续依靠美国信息资讯频道为其提供准确的信息。如果出现气象预报员与公众之间互不信任的状况，即使在最该听信预报员的话的时候，公众也不会这样做。

尽可能地做出准确预测

正如马克斯·梅菲尔德告知美国国会的那样，他活了 60 多年，对像卡特里娜飓风袭击新奥尔良这样的灾难早已作好了心理准备。梅菲尔德在恶劣的天气条件中长大，他从小生活在俄克拉荷马州，那里是龙卷风多发地，他的预测生涯始于空军部队，那里的人们都认真对待风险，并且会依此制定作战计划。对于美国国家飓风中心来说，向普通民众传达预测实属不易，梅菲尔德花了很长时间才意识到这一点。

梅菲尔德用他那俄克拉荷马州人特有的慢悠悠的语气说："1989 年吉尔伯特飓风发生后，我与一位来自佛罗里达州的行为科学家聊天，他说人们对飓风警报无动于衷。当时我觉得自己受到了侮辱，认为人们当然注意到了飓风警报。但我也知道这位科学家的说法完全正确。人们只是对'飓风警报'这个词无动于衷，对当地官员的说法他们还是比较关注的。毕竟，你不能要求预测者或电视主持人去决定何时开放避难所，或何时改变航道。"

在梅菲尔德的指导下，美国国家飓风中心开始更多地关注应该如何向公众公布预测信息。大多数政府机构的网页多年都不曾更新，看起来总是和你的邮筒中还能收到"美国在线"免费光盘的时候一样。而飓风中心则不然，它在产品设计上非常用心，制作出一系列色彩丰富、引人注目的图，直观、准确地传达了信息，涉及内容多样，从风速到风暴潮无一不包。

美国国家飓风中心也非常谨慎地呈现预测中的不确定性。梅菲尔德说："不确定性是气象预报的基本要素，没有对不确定性进行描述的预测是不完整的。"比如，美国国家飓风中心的预测图中不会只显示一条单一的飓风行进轨迹，而是会突出描绘一个锥形的不确定范围，有人称为"锥形混沌"，它能显示出"飓风眼"最有可能登陆的区域。即使这样，梅菲尔德仍然担心做得不够。像山洪暴发这样的重大灾难，其位置可以与暴风雨中心相距很远，并且在疾风平息之后才出现，所以山洪暴发的后果通常比引起山洪暴发的暴风雨本身还严重。2011 年，大众媒体对暴风雨进行了大肆宣传，纽约市没有一人在艾琳飓风中丧命，而在电视台摄像师结束报道时，内陆地区佛蒙特州却有 3 人死于飓风之后泛滥的洪水。

美国国家飓风中心通常不会向地方政府下达"对城市居民进行疏散"这类指导政策，而是把疏散任务分配给美国国家气象局下属的 122 个地方办公室，由地方办公室与州长、市长、治安官和警官进行沟通。对此，官方给出的理由是，美国国家飓风中心认为当地办公室对当地文化和居民更为了解，工作起来比较得心应手。而在我与梅菲尔德交谈之后，得到的非官方理由是，美国国家飓风中心想要与美国国家气象局划

清界限。美国国家飓风中心希望其独立发布的飓风预测能够尽可能地准确、真实，并且不受任何干扰。

但这种分离方法在新奥尔良却没有奏效，梅菲尔德不得不自己处理。

做出疏散的决定并非易事，一部分原因是疏散本身就存在危险性。2005 年“丽塔”飓风时，一辆载有多名被疏散者的大巴正开往医院，在离开休斯敦时突然起火，车上 23 名老年乘客不幸罹难。梅菲尔德说：“对于当地的管理者来说，这是相当棘手的问题，他们看到的只是概率性的信息，却必须依此作决定，一旦决定了，便没有其他选择，非此即彼。

然而，若是这样就明显需要疏散，而疏散信息却没有得到明确传达。

梅菲尔德有一次说起：“美国国家飓风中心有一个叫马修·格林的小伙子，是个特别出色的年轻人，拥有气象学的学位，负责向从事运输工作的人发出飓风预警。马修·格林的母亲住在新奥尔良，出于某种原因，飓风到来时她没有离开。看看这个家伙，他了解飓风，也懂得应急管理，却没有办法让自己的妈妈撤离。”

于是，美国国家飓风中心开始不停地打电话给墨西哥湾岸区的地方官员。2005 年 8 月 27 日，星期六——此时气象预测表明情况会更加糟糕，但距离卡特里娜飓风来袭还有两天时间——梅菲尔德找到密西西比州州长黑利·巴伯和路易斯安那州州长凯瑟琳·布兰科，前者几乎马上就向本州最易受灾地区下达了强制疏散的命令，后者则宣布本州进入紧急状态。布兰科请梅菲尔德给新奥尔良市长雷·纳金打个电话，告诉他尽快做好预防工作，因为这位市长的反应总是慢半拍。

纳金没有接到梅菲尔德的电话，但是他给梅菲尔德回拨了电话。梅菲尔德对我说：“我记不清纳金当时具体说了什么，我们当时两三天就有一大堆的会谈，但是我清楚地记得我告诉过他，‘你需要作一些艰难的决定，有可能会面临大量的人员伤亡。’”梅菲尔德让纳金下达强制撤离命令，越快越好。

纳金没把梅菲尔德的话当回事儿，只是下达了自愿撤离的命令。在《越轨干探》这部电影中，这个命令就意味着“放轻松”，只有强制撤离的命令才能传达出这次威胁的强烈程度。1965 年新奥尔良遭遇贝齐飓风的袭击，大多数新奥尔良人在那次大灾难中丧生，按照定义，那些活着的人就算是飓风灾难的幸存者了。后来，一位新奥尔良的老市民告诉政府官员：“既然我能在贝齐飓风中幸免于难，我也能在这次飓风中逃过一劫，我们都治得了飓风，放心吧。”这种反应见惯不怪。从对卡特里娜飓风和其他暴风雨的调查中不难发现，经历过一次飓风袭击的人都

不大可能在下一次飓风来临时撤离。

纳金推迟下达撤离命令是因为害怕纠纷——如果因此搅了旅馆的生意，那么旅馆经营者可能会控告他。总之，直到 2005 年 8 月 28 日上午 11 点，纳金才下达“强制撤离”的命令，而那时仍未得知这一命令的市民都不知所措。一项调查显示，拒绝撤离的人中约有 1/3 的人是因为根本没有接到撤离命令；另有 1/3 的市民接到了命令，却没有从中得到明确指示。对灾难受害者的调查并非总是可信的——在情绪高度紧张时，人们说不清楚自己当时为什么那样做，即使命令早已下达并重复再三，还是有一小部分市民会说他们从来没有接到疏散命令。但在这种情况下，纳金要为这种混乱局面负主要责任。

当然，对于卡特里娜飓风，人们仍是怨声载道，除了对纳金多有责难，他们对美国联邦应急管理局（FEMA）也同样心存不满。当然也有值得表扬的地方，大多数市民确实撤离了，部分原因在于飓风中心的准确预测。如果 1965 年贝齐飓风冲破了堤顶，那根本不用等到可靠的预测出现，死亡人数就很有可能会远远超过卡特里娜飓风造成的死亡人数。

然而，卡特里娜飓风带给人们的教训是，准确性才是预测者的第一要务；而认为政治、个人荣誉或经济利益高于预测的真实性，其实是预测的原罪。有时这样做的意图是好的，但总会使预测更糟。美国国家飓风中心竭尽所能地避免预测向利益妥协。与本书中所有预测失败的例子形成对照的是，仅在过去 25 年里的时间里，飓风中心的预测准确率就提高了 350%，这绝非偶然。

梅菲尔德说：“预测者的任务，是尽可能地做出准确预测。”说起来简单，可相当多领域的预测者通常还是会出现预测错误。

[1]不幸的是，尽管预测人员对飓风登陆地点的预测已经算是准确了，但他们对其着陆时强度的预测并不是很准确。究其原因，是因为控制风暴强度的风力远小于决定其路线的风力，这就意味着他们需要更出色的点阵，这一点即使蓝火计算机也无法做到。

第五章

地震预测：一个困惑了人类 1 000 年的难题

2009 年 4 月一个寒冷的星期天晚上，意大利拉奎拉的居民们正准备入睡，突然感觉到一阵阵晃动，感觉不是太强烈，每一次震动都像是一列货运火车从远处经过，其实是发生了地震。第一次地震发生在当地时间夜里 11 点，震级为 3.9 级。第二次地震的震级更低些，矩震级为 3.5 级（下文省略矩震级），都不足以震醒熟睡的人。

坐落在亚平宁山脉上的拉奎拉的确处于地震带，这个以滑雪胜地和中世纪城墙闻名的旅游小镇最近经历了一系列的地震，数量多得不正常，星期天晚上发生的两次地震是那一周里发生的第 7 次和第 8 次 3 级以上的地震。在这个地区，小级别的地震不算稀奇，但频率没有这样高，通常两三个月才发生一次，这次的频率几乎是平时的 100 倍。

与此同时，在亚平宁山脉的尽头，苏尔莫纳镇的居民刚刚从地震的恐慌中缓过来。意大利国家原子物理研究所的詹保罗·朱利安尼称，他测量出该地区的氡气含量异常的高，由此推断这是地震的先兆，马上跑去告诉苏尔莫纳的镇长，2009 年 3 月 29 日下午苏尔莫纳镇会发生地震。镇长被说服了，派了几辆货车带着扩音器绕着小镇到处广播，警告居民要发生地震了。

可是，3 月 29 日那天苏尔莫纳镇并没有发生地震，詹保罗·朱利安尼的预测失败了。有人向相关部门控告他，说他是在制造恐慌。当地政府勒令他撤下公布在网站上的预测，以免造成更大的骚乱。

在发生了地震的拉奎拉，当地政府却告诉民众不必担心地震群。意大利民事保护部副部长伯纳多·德博纳迪尼斯解释说，地质断层带活跃有助于释放能量，减少大地震的威胁。德博纳迪尼斯同意某位记者的观点，认为民众们可以放松警惕，喝杯小酒，他还向大家推荐当地一款名为蒙特普恰诺的特色酒。

然而，星期一凌晨，当地时间 3 点 32 分，一场 6.3 级的大地震袭击了拉奎拉。房子倾斜，屋顶塌陷，家具散架，300 多人丧生，65 000 人无家可归，经济损失超过 160 亿美元。

地震可以预测吗？

拉奎拉本应更好地应对这场地震。这座城市靠近一个十分活跃的地质断层带，也称潜没带（指地壳的板块沉到另一板块之下），此断层带是由覆盖在地球表面的八大板块之一的非洲板块缓慢、持久地挤向亚欧

板块下方形成的。在拉奎拉，第一次有记载的地震发生在 1315 年，随后在 1349 年、1452 年、1461 年、1501 年、1646 年、1703 年和 1706 年都发生过地震，其中最严重的一次地震发生在 1783 年，当时超过 5 000 人丧生。但每次地震后，教皇都会下令重建小镇，人口再度繁衍。

此后两个多世纪，拉奎拉这个小镇仍冒险地留在原地，备受地震威胁。1958 年拉奎拉发生了一场 5.0 级的地震，震级不算太高，镇里上了年纪的人对此还有些印象。但 2009 年这场地震的强度要大得多。我们知道，震级与释放的能量呈对数关系，震级每增加一个点，释放的能量会增强 32 倍。所以，2009 年那场 6.3 级的地震释放的能量几乎比 1958 年的那场地震多出 75 倍，比强震前那个晚上居民感受到的小震颤——强震的震前波——多出 3 000 多倍。

在意大利，6.3 级地震已是震级非常强的地震了，但从全球范围来看，这只不过是地球打了个嗝而已。2011 年摧毁日本的那场大地震高达 9.0 级甚至 9.1 级，释放的能量是 6.3 级地震的 11 000 倍。自有可靠的地震测量开始，所记载的震级最高的地震为 1960 年发生在智利的 9.5 级地震，释放的能量约是 6.3 级地震的 60 000 倍。

既然 6.3 级的震级并不算高，为什么在意大利这样一个高度工业化的富裕国家，富有的拉奎拉小镇却会遭受如此大的损失呢？原因之一是这座城市的地质环境。拉奎拉坐落在一个古老的湖床上，这就决定了它很容易受到地震的影响。墨西哥城同样建立在古老的湖床上，1985 年的一场地震夺去了 10 000 人的生命，而震源距地面超过 322 公里，由此可见湖床的脆弱。

主要原因还是由于拉奎拉小镇的麻痹、疏忽，对来自地下 15 公里处的地震危险毫不在意，也没有任何应对准备，比如建筑物标准的规范，应急物资的准备以及社区抗震演习等。结果，不仅那些有着几百年历史的古老建筑在地震中受损，许多现代建筑，包括某家医院在 2000 年刚盖的侧楼也未能幸免。当初哪怕只是稍稍提醒一下市民，都有可能挽救那些逝去的生命。

詹保罗·朱利安尼是否预测到拉奎拉的地震了呢？在意大利，许多小报将朱利安尼描绘成圣人或殉道者，他的声音轻柔、头发蓬松，经常穿着和当地足球队队服颜色一样的衣服，他扮演的角色也许是一个谦虚的公务员，也许是一个看上去心不在焉、观点常被科研机构忽略的教授。朱利安尼称已经警告过家人和朋友拉奎拉镇会发生地震，但是警方禁止他向其他人发出这个警告。他要求拉奎拉政府就此事道歉，不是向他道歉，而是向拉奎拉的民众道歉。

暂且不论朱利安尼是否真的预测到拉奎拉的地震，他发布的地震预

测非常明确：苏尔莫纳镇（而非拉奎拉）面临较大风险，另外，他还认为地震会在 2009 年 3 月发生，并非 4 月。事实上，朱利安尼曾对当地的一家报纸说危险已经过去了。在开始长篇大论地解释潮汐的影响之前，朱利安尼说：“这些概念，简单地说就是地月系统自转到了近日点……在距离地球最近的位置，和金星排成一行……总之，我可以向我的同胞保证，地震群将在 3 月底逐渐减少。”

近日点、金星、氡气含量，这些和预测地震有关系吗？与朱利安尼错误地预测苏尔莫纳地震，又有什么关系？这些都不重要。当灾难来袭时，我们会从噪声中寻找信号，寻找有助于解释周围混乱状态的所有信息，使世界恢复秩序。朱利安尼杂乱无章的解释就是这一现象最生动的阐释。

在各类灾难中，地震对我们的秩序感干扰最为严重，它简直可以撼动我们的根基。同样是自然灾害，飓风从天而降，常和“上帝的眷顾”联系在一起，但从地球深处爆发的地震却常被视为“上帝的盛怒”，无情而虚幻。（1755 年的里斯本地震推动了现世哲学的发展。）另外，飓风及随之而来的洪水、龙卷风，还有火山喷发通常都能提前预测，唯有地震例外，人们努力了几个世纪，却还是很难对其进行预测。

“圣杯”根本就不存在

很长一段时间以来，加利福尼亚州的帕萨迪纳市都是全球地震研究的中心，美国加州理工学院就坐落在这里，1935 年，查尔斯·里克特在这里创立了著名的震级对数函数。美国地质调查局在此也设有办事处，其大部分地震专家都常驻此地。2009 年 9 月，我前往帕萨迪纳市拜访苏珊·休博士，休是美国地质调查局的一名地质专家，写过很多本关于地震预测的书。她曾经看过朱利安尼的电视采访，对他的说法很是怀疑，并在《纽约时报》上写了一篇言辞激烈的社论，对朱利安尼本人和那些关注朱利安尼的人都提出了批评。

休的这篇社论认为朱利安尼的成功不过是个偶然。休说：“民众之所以听说了朱利安尼先生的预测，是因为这个预测貌似经过了证实。但是，还有很多民众没有听说过的‘不正确的’预测。”

如果你让上百人设法做出地震预测，而每年又有数百场地震发生，那必然有人会预测对一次。资深的地质学家已经多次证实，朱利安尼的氡气和月运周期理论对地震预测的作用不大，或者根本一点儿用处都没有。朱利安尼的运气一直不错，就好比一只能用键盘敲出莎士比亚作品的猴子，或者一只会预测世界杯比赛结果的章鱼。

休在美国地质调查局的办公室位于加州理工学院一个安静的角落附

近，周围桉树环绕，少有学生经过。我们见面时，她刚结束一项在土耳其研究地质断层系统的工作，看上去风尘仆仆。慈眉善目的休，一头卷发，黑色的眼睛，疲惫中带着怀疑。打过招呼后不久她就向我发问："你的日常工作是什么？"

交谈中，休不时地拿起办公桌上放置的一个口袋形状大小的地球仪，像是从机场的礼品店里买的，然后用食指在上面画出一条线，从日本海向东和东南方向延伸。

提及全球破坏力最大的地震时，休向我讲解道："这些地震真的都集中发生在这一地带——从中国南部一直延伸到希腊。这个地震带十分复杂，该地区的许多建筑物又都不够结实，如果德黑兰发生一场特大地震，估计会有上百万人丧生。"

的确，在现代史上，几乎所有最致命的地震都发生在休划定的这一地带，穿越中东文明的摇篮和地球上人口最稠密的地区，包括中国和印度。这些地区通常贫穷而拥挤，缺乏足够的物资应对 300 年一遇的灾难，地震发生时，死亡人数常会达到几十万。

实际上，尽管地震看似比飓风罕见，但造成的死亡人数却比飓风多得多。也许这是因为地震很难准确预测到，而飓风却不同。目前，飓风登陆地点的预测准确率已是 25 年前的 3 倍，自从公元 9 世纪日本第一个宣称通过观察鲶鱼的行为能预测地震以来，这么长久的时间过去了，地震预测仍没有什么大的科技方面的进展。（据多次报道称，奶牛、猪、鳗鱼、老鼠、长尾小鹦鹉、海鸥、乌龟、金鱼、蛇等动物在震前都出现过反常行为。）

不仅意大利的小报会关注像朱利安尼这类古怪的人的观点，很多普通人也都把他们的话当真了。美国加州地震预测委员会每年都会收到数百份未经证实的地震预测，该委员会称，其中大多数预测无非是在"讨论家里宠物的奇怪行为，或是个人直觉，或是家中亲属关节的疼痛症状，还有一些神秘的信号或预兆是科学家根本无法理解的。"同时，就连学术期刊上的一些预测都与日本古老的民间传说没什么两样。比如，2010 年发表在颇有声望的期刊《动物学杂志》上的一篇论文，观察到在拉奎拉地震前 5 天里，距地震点约 80 公里的池塘里，蟾蜍全部都停止产卵。让人瞠目的是，文章称这就是作者预测到地震的证据。



图 **5–1** 自 **1900** 年以来最致命的地震

资料来源：美国地质调查局

就是诸如此类的研究让休精疲力竭。休告诉我："如果时光可以倒流，你肯定愿意回到 20 世纪 70 年代，那时人们会有一些想法——他们总是踌躇满志的——但 10 年后，该方法会暴露出缺点，又一个 10 年过去了，你又有了新的方法，可再过 10 年，这个方法的弊端又显露出来。听到这里，你大概明白了，多数优秀的地震科学家都很清楚这一点，他们不会去追寻那个可能根本就不存在的'圣杯'。"

但是，真的没有办法预测地震了吗？朱利安尼的金星理论和蟾蜍预兆可以很容易被排除，那在拉奎拉大地震发生之前，周围地区的小震群可以用来进行预测吗，还是这些只是巧合？众所周知，地震学界是出了名的保守，例如地震学界花了很长时间才接受板块构造理论，现在大家都同意大陆板块的碰撞是引发地震的首要原因，而早在 1912 年这一理论已被提出，但直到 20 世纪 60 年代才被采纳。那么，休表现出的怀疑精神是否有些过头，变成了愤世嫉俗的态度？

美国地质调查局的官方立场更加明确：地震就是不可预测的。该机构在其官网上声称，“无论是美国地质调查局，还是美国加州理工学院，或是任何一个科学家都没有预测出大地震。因为大家都不知道该如何预测，也不想知道要过多久才能实现？”

地震真的不能预测吗？本书是关于预测的，而不是关于如何进行预测的，但我愿铤而走险，作个预测：2013 年日本发生的地震次数要比新泽西多；22 世纪的某个时刻，会有一场大地震袭击加利福尼亚州。

其实，我和美国地质调查局都是在玩文字游戏。不同的领域，“预测”和“预言”所指也有所不同；在一些学科中，它们可以相互替换，但在其他学科领域，它们则存在着区别。这二者区分最明显的当属地质学，地质学家认为：

1. 预言是对何时何地会发生地震的明确、具体的陈述，比如，6 月 28 日，日本京都将发生一场大地震。
2. 预测针对的时限较长，是对地震发生的可能性的表述，比如，未来 30 年，南加利福尼亚州有 60%的概率会发生地震。

美国地质调查局的官方观点是，地震不能被预言，但是，地震可以被预测。

我们对于地震的了解

访问美国地质调查局的网站时，你会发现该网站提供了很多帮助你预测地震的工具，其中有一个特别简便的应用程序，只要输入美国任意一个地方的经纬度，这个应用程序就会预测出一段时间内该地发生地震的概率。根据这项应用，我在表 5–1 中列出了美国地质调查局网站提供的美国不同地区主要城市发生地震的可能性。

众所周知，美国加州属于地震频发区，美国地质调查局推算出旧金山每 35 年就会发生一次震级为 6.8 级或更大级别的地震。另外，阿拉斯加州的地震也很频繁，历史记载的第二大地震，就是 1964 年发生在安克雷奇的 9.4 级地震。

知道南加州的查尔斯顿市吗？这座城市的地震活动也很频繁，1886 年该市就经历了一场震级为 7.3 级的大地震。美国地质调查局预计，该市发生大地震的频率约为 600 年一次。在西雅图，估计人们得时刻准备应对地震，这里的地震频率比加州其他城市都高。但是，在丹佛的居民就不必担心地震，这里离任何一个大陆板块交界处都有好长一段距离。

表 **5–1** 美国部分城市方圆 **80** 公里范围大地震（＞ **=6.75** 级）发生频率

安克雷奇	每 30 年 1 次
旧金山	每 35 年 1 次
洛杉矶	每 40 年 1 次
西雅图	每 150 年 1 次
萨克拉曼多	每 180 年 1 次
圣迭戈	每 190 年 1 次
盐湖城	每 200 年 1 次
波特兰	每 500 年 1 次
查尔斯顿	每 600 年 1 次
拉斯韦加斯	每 1 200 年 1 次
孟菲斯	每 2 500 年 1 次
菲尼克斯	每 7 500 年 1 次
纽约	每 12 000 年 1 次
波士顿	每 15 000 年 1 次
费城	每 17 000 年 1 次
圣路易斯	每 23 000 年 1 次
亚特兰大	每 30 000 年 1 次
丹佛	每 40 000 年 1 次
华盛顿	每 55 000 年 1 次
芝加哥	每 75 000 年 1 次
休斯敦	每 100 000 年 1 次

达拉斯	每 130 000 年 1 次
迈阿密	每 140 000 年 1 次

对于一个认为地震是无法预测的机构而言，这些具体而又容易获取的信息似乎太多了。美国地质调查局使用的是广为流传的古登堡–里克特法则，这一法则是由查尔斯·里克特和他在加州理工学院的同事贝诺·古登堡仔细研究过地震数据后，于 1944 年创立的。该法则认为，地震级别和发生频率之间的关系是相对简单的。

如果将地震的频率和震级进行比对，你会发现震级升高时，地震发生频率则会大大降低。世界上发生的灾难性地震屈指可数，但是小地震却多达数百万次——2.0~2.9 级的地震每年都约发生为 130 万次，当然，对其中大多数的地震，人们并未察觉到，有时连地震仪也检测不到。然而，今天的地震无论发生在哪里，只要震级大于或等于 4.5 级，几乎都会记录在案。图 5–2A 依据 1964 年 1 月~2012 年 3 月的实际地震记录而来，显示出地震发生频率的急速下降趋势。

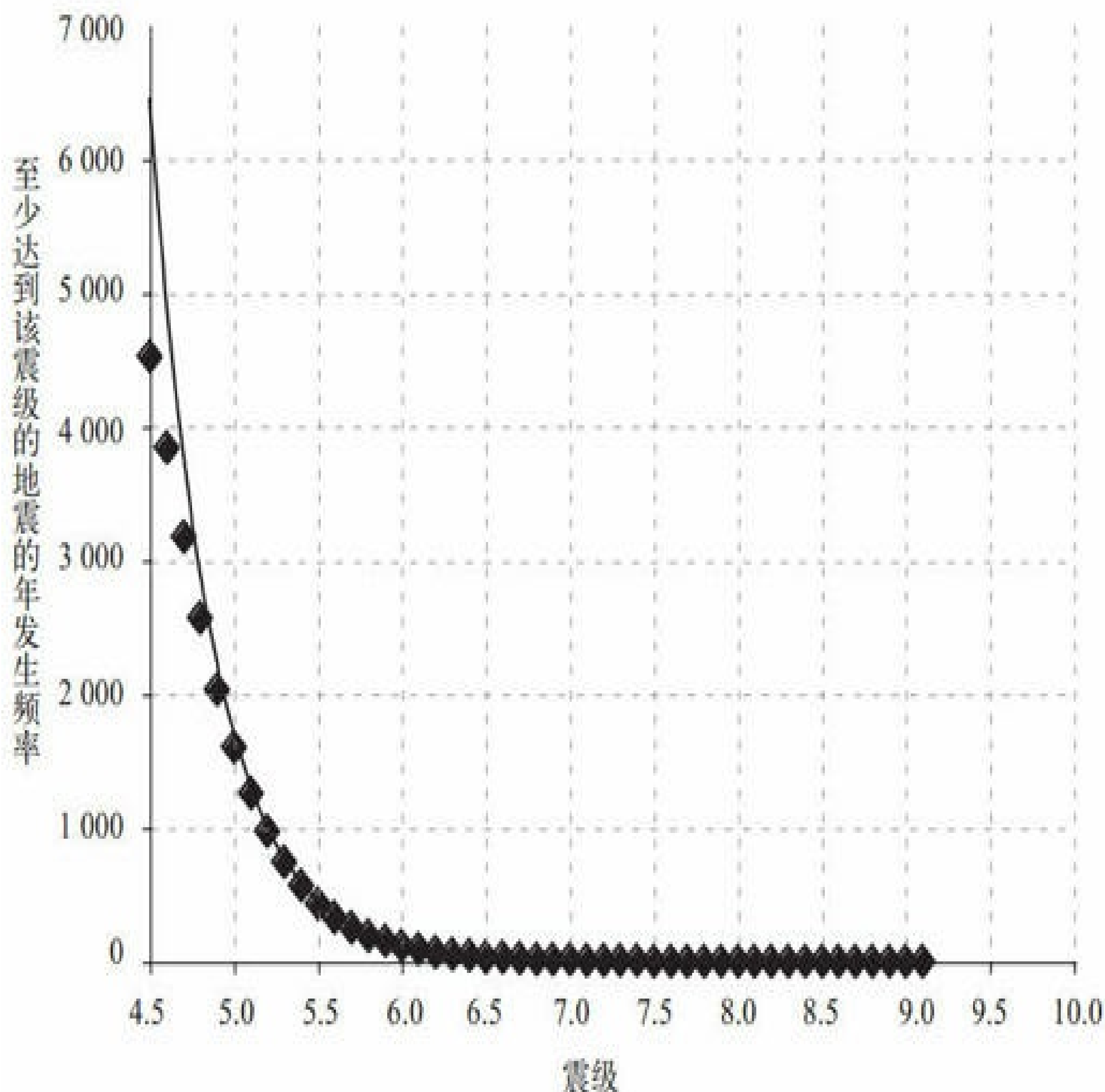


图 **5–2A** 世界范围内地震发生频率（**1964** 年 **1** 月 **~2012** 年 **3** 月）

对地震曲线图稍加改变，这些地震就会呈现出惊人的规律性。在图 5–2B 中，我把显示“不同级别地震发生频率”的纵轴换成对数函数数轴，此时图中呈现的地震频率完全成了一条直线，这个图形具有幂律分布的特征，幂次法则也是里克特和古登堡发现的。

遵循幂律分布的事物具有较高的使用价值：你可以根据小级别地震发生的次数预测大地震发生的频率，反之亦然。事实证明，震级每提高一个点，地震发生的频率就会降低 10 倍，所以，6 级地震发生的频率是 7 级地震发生频率的 10 倍，是 8 级地震发生频率的 100 倍。

在全世界，或者说在地球范围内，古登堡–里克特法则都是适用的。

例如，假设我们要为伊朗首都德黑兰进行地震预测。当地自有地震测量开始就没有发生过灾难性地震，但德黑兰周围经常发生中等级别的地震，1960~2009 年的近 50 年时间里，该地区共发生了 15 次震级达 5.0~5.9 级的地震，大概每 3 年发生一次。根据古登堡和里克特发现的幂次法则，这就意味着，德黑兰发生震级为 6.0~6.9 级地震的频率为每 30 年一次，7.0 级或更高级别地震的发生频率是每 300 年一次。苏珊·休担忧的正是这样的大地震。

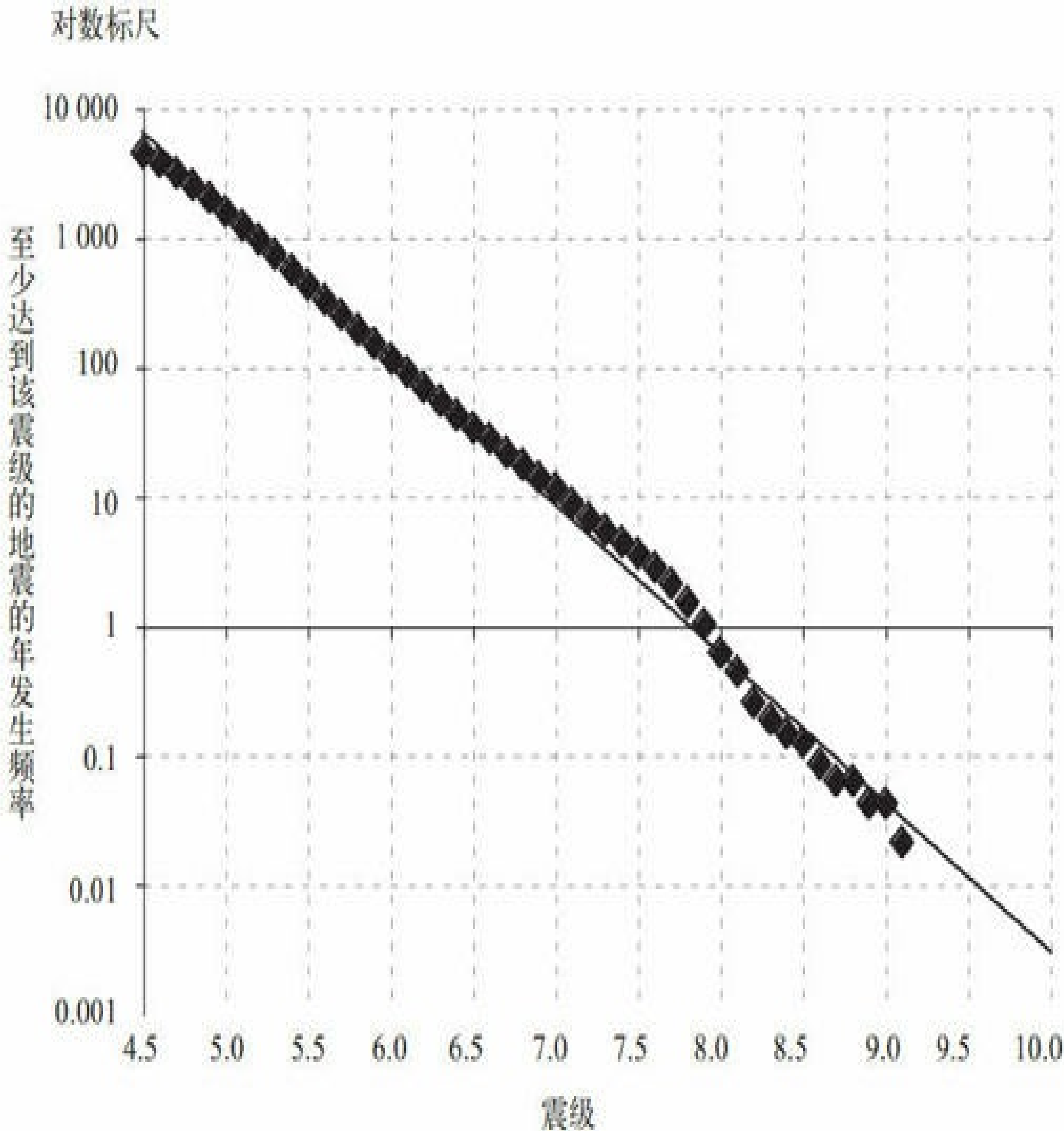


图 **5–2B** 世界范围内地震发生频率（**1964** 年 **1** 月 **~2012** 年 **3** 月）

2010 年海地发生 7.0 级大地震，死亡人数达 316 000 人，表明地震

可能对发展中国家造成灭顶之灾。海地的很多问题在伊朗同样存在，国家贫困、政治腐败、建筑物抗震标准宽松，只是伊朗的人口比海地稠密得多。依据伊朗以往震级较低就会造成高伤亡的情况看，美国地质调查局预计，一旦发生灾难性地震，德黑兰的人口会减少 15%~30%，也就是一个人口为 1 300 万的城市会有 200 万~400 万人丧生。

然而，古登堡–里克特法则并没有告诉我们，地震会在什么时候发生（这一法则也没有指明，如果德黑兰近期没有发生地震，未来必定会发生）。像伊朗和海地这样的国家，根本就没有物质能力来应对 300 年一遇的灾难。确实，依据古登堡–里克特法则做出的地震预测，可以为一个地区提供周全的总体指导。但是，就像气象预报一样，单纯依靠数据记录做出的预测无法直接转换成我们的行动指南（比如，预报说 3 月份伦敦有 35%的时间在下雨，但这条信息无法告诉我们出门是否应该带伞）。此外，地质活动的时间范围长达几百年乃至一千年，但人类的寿命却只有数十年。

究竟是信号还是噪声？

地震学家真正感兴趣的是含时预测，苏珊·休称之为地震学的“圣杯”。在含时预测中，地震发生的概率在一段时间内总会出现波动。

然而，即便是那些根本就不相信可以进行含时预测的地震学家也承认，地震的分布是有模式可循的，最明显的一点就是余震的存在。几乎所有大地震之后都会发生几十次甚至几千次余震（2011 年日本大地震后就发生至少 1 200 次余震），这些余震通常遵循一个可预测的模式，一场地震发生后，余震更有可能会随即发生，而不是几天之后才发生，几周之后发生的概率就更低了。

然而，这种预测模式对于挽救生命并没有实质性帮助，因为根据定义，余震总是比最初的主震威力小。如果一个地震带发生了一场威力足够大的地震，通常会伴随许多小余震，然后地震会暂且告一段落。然而，实际情况往往不是这样。比如，1811 年 12 月 16 日在密苏里州和田纳西州交界处的新马德里地震带发生了一场大地震，据地震学家评估，此次地震的震级为 8.2 级，6 个小时之后，竟然又发生了一场震级为 8.2 级的地震。而该地震带仍然没有平静下来，继 12 月 16 日的两次大地震之后，1812 年 1 月 23 日又发生了一场震级为 8.1 级的大地震，2 月 7 日的地震的震级更是达到了 8.3 级。这几次地震中究竟哪些是前震？哪些又是余震？所有的解释似乎都说不通。

当然，问题的关键在于，我们能否在地震发生之前就预测到，能否提前分清前震和余震。当我们跨越时间和空间观察地震分布数据时，这

些数据似乎在诱使我们相信，在噪声中可能会找到信号。

比如，图 5–3A 显示了拉奎拉周边的地震统计数据，从 2006 年起到 2009 年发生那次 6.3 级地震为止。图中除了一个代表主震的黑色大圆圈外，剩下的数据点都代表主震之前的每次小地震。在拉奎拉的例子中，似乎确有一个可辨识的模式。2009 年大地震前发生了一连串的地震，总的震级约为 4 级，比该地区以往的地震活动都频繁。

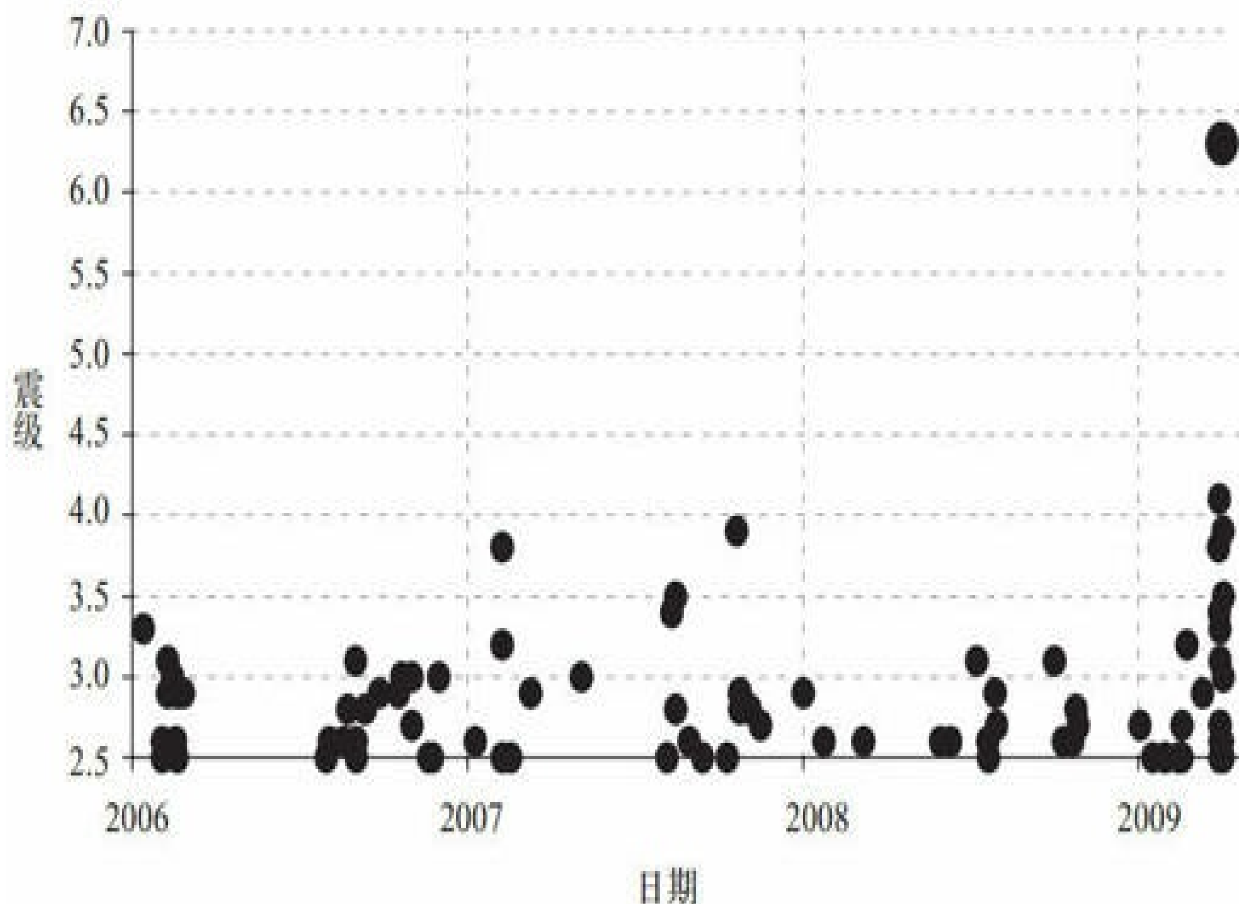


图 **5–3A** 意大利拉奎拉周边地区地震统计（**2006** 年 **1** 月 **1** 日 **~2009** 年 **4** 月 **6** 日）

另一个更有争议的例子是 2011 年发生在日本的大地震。在为日本东北部地区绘制地震分布图时（图 5–3B），我们的第一印象是，这一地区的地震活动比意大利活跃很多。该地区地震发生的时间是否有模式可循呢？答案似乎是肯定的。例如，2008 年中期，该地区发生了一系列震级为 5.5~7.0 级的地震，此后并没有发生更大的地震。但我们的确看到，2011 年 3 月 9 日发生了一场格外剧烈的前震，震级为 7.5 级，55 个小时后日本东北部发生了 9.1 级的大地震。

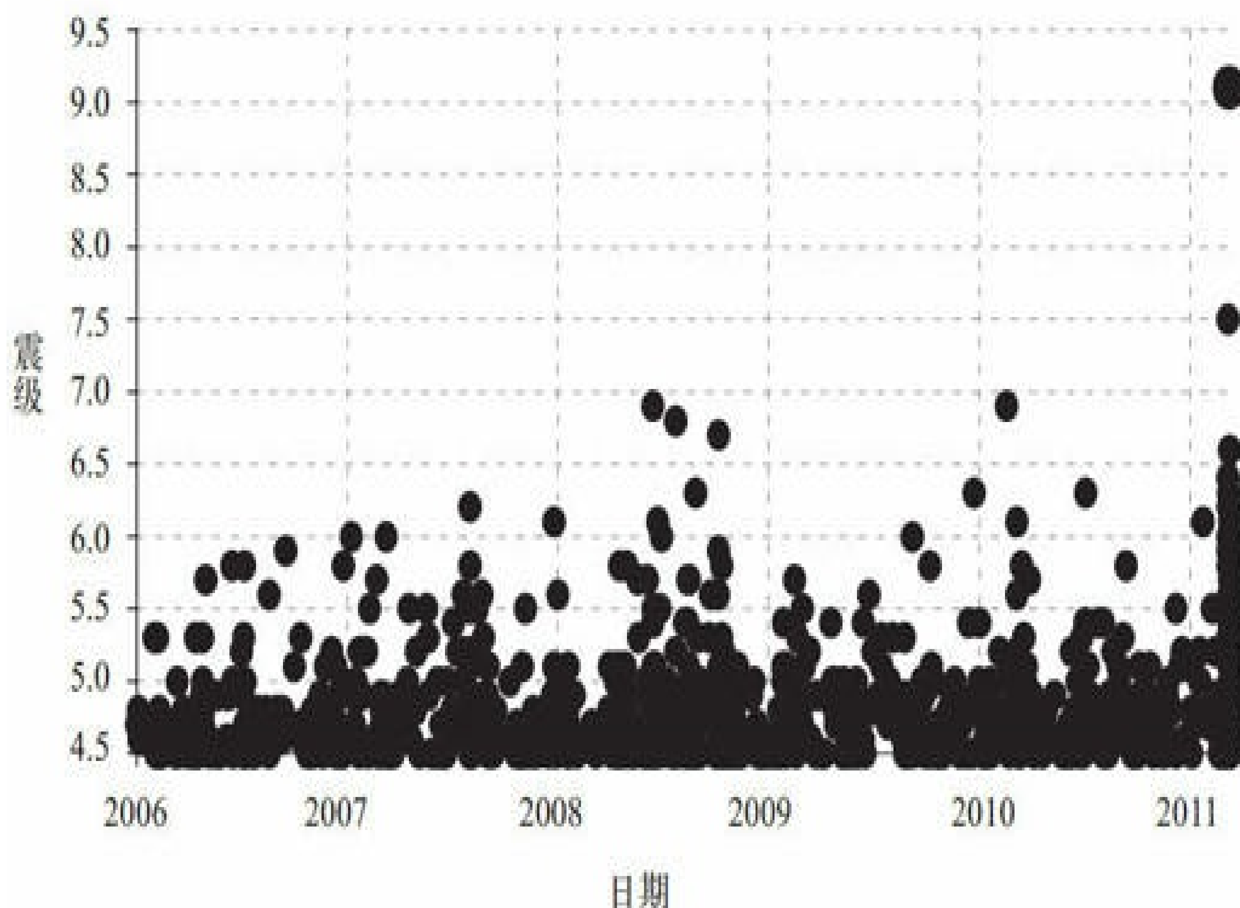


图 **5–3B** 日本东北部地区地震统计（**2006** 年 **1** 月 **1** 日 **~2011** 年 **3** 月 **11** 日）

然而，只有一半地震有可辨识的前震，海地那次地震就没有前震（见图 5–3C）。在加勒比海的大部分地区，测量地震的仪器都相对落后，所以没有发生震级为 2~3 级地震的记录，但是美国或是其他地区的地震仪应该能够测量到所有记录在案的 4 级及以上级别的地震。该地区的最近一次 4 级地震发生在 2005 年，震级为 7.0 级的海地地震发生在 5 年后。也就说，这次记录并无预警作用。

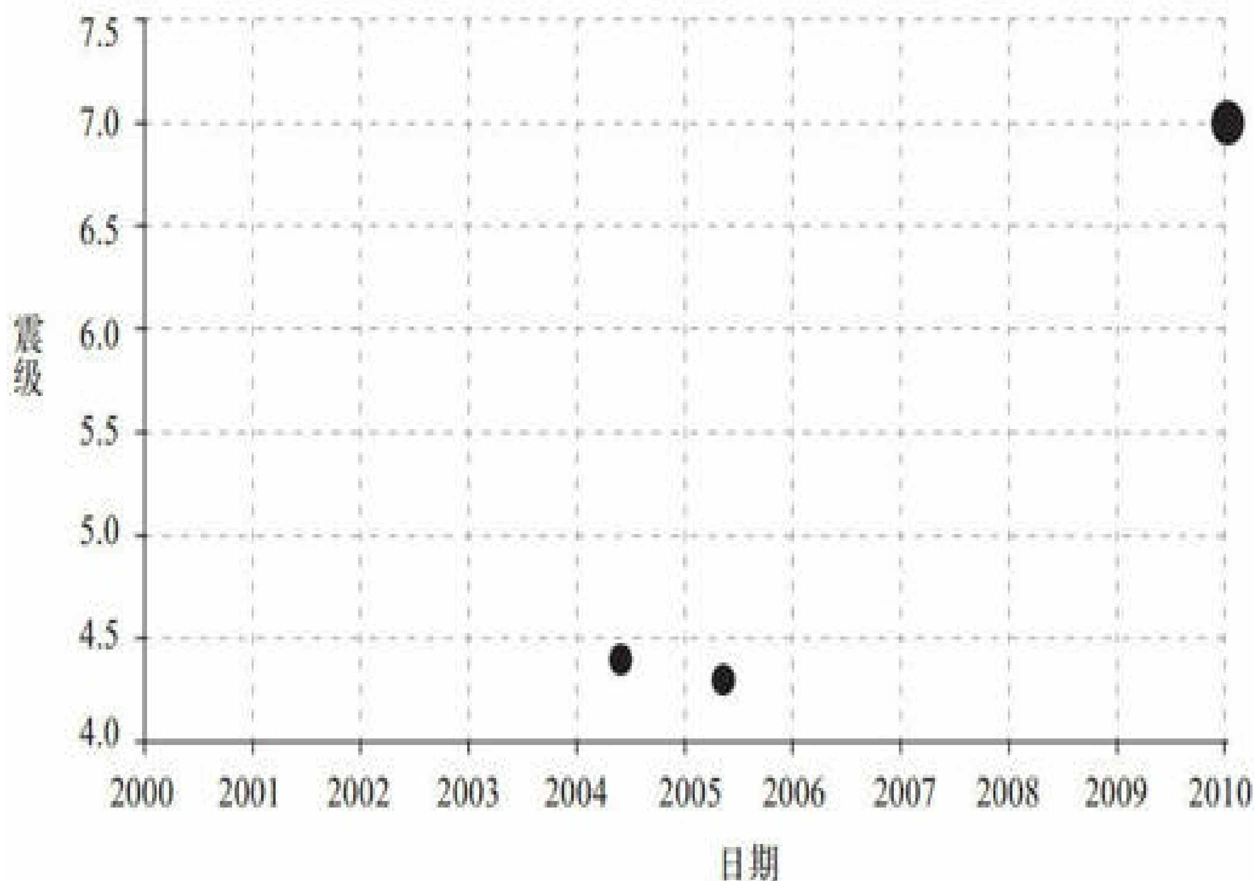


图 **5–3C** 海地莱奥甘周边地区地震统计（**2000** 年 **1** 月 **1** 日 **~2010** 年 **1** 月 **12** 日）

更复杂的情况是错误的预警，误认为地震活动高发期是大地震的预警，结果却往往没有发生大地震。地震学家熟知的一个例子就是，2008 年年初发生在内华达州里诺市的一系列小地震。里诺的地震群和 2009 年拉奎拉附近的地震群很相似，但结果却不一样。里诺的地震群中，最大的一次仅为 5.0 级，之后就没有更大的地震发生了。

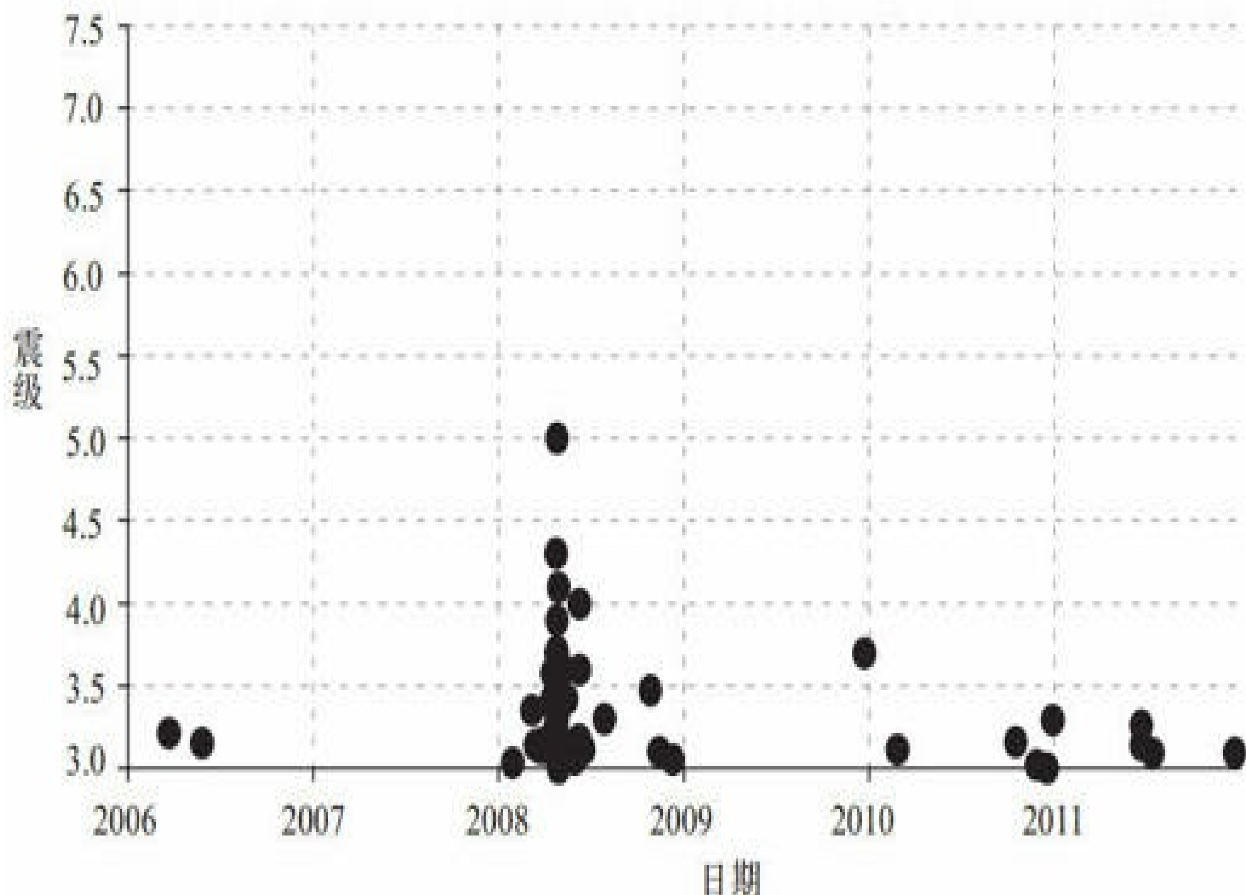


图 5–3D 内华达州里诺周边地区地震统计（2006 年 1 月 1 日 ~2011 年 12 月 31 日）

以上的数据让地震学家们感觉很难处理，但这只是冰山一角。数据似乎总处于一种让人陷入炼狱的状态，不是那么无规律，也不是那么可预测的。但这至少表明我们已经在探索的路上，并在地震预测方面取得了一些进步。但是，史上记载的关于地震预测的尝试，几乎全都是失败的。

那些以失败收场的地震预测

2009 年，休出版了一本书，名为《预测不可预测之物：地震预测科学的混乱现状》，这是一本记载地震预测历程的书，它对预测行业的批判就像菲尔·特罗克对政治评论家的批判。因为在地震预测行业中，进步似乎太少，而错误警告却太多。

秘鲁，利马

在比较知名的案例中，有一例与地球物理学家布莱恩·布雷迪有关，这位地球物理学家拥有麻省理工学院博士学位，现任职于科罗拉多州矿业大学。布雷迪断言，1981 年秘鲁首都利马会发生一场震级为 9.2 级地

震，堪称有史记载的最高震级的地震之一。布雷迪的预测最初得到了地震学界的强力支持，他和一位美国地质调查局的科学家共同完成了这一预测的早期版本。但后来，他的理论变得太过复杂，最终这一理论可谓无所不包，从在研究中观察到的岩层突裂跨越到爱因斯坦相对论，认为相对论可以支持岩爆现象。同事们告诉布雷迪，这一理论已经超出了他们的理解能力，其实同事们是在委婉地告诉布雷迪，“你是个怪人”。1981 年 8 月，布雷迪最后预测说，秘鲁会发生一连串地震，9.2 级地震只是其中一次，最强震级会达到史上最大的 9.9 级。

结果，有人将这一预测结果泄露给秘鲁媒体，引起了民众的恐慌，而这个看上去严肃认真的美国科学家似乎十分肯定秘鲁的首都将会变成一片废墟。秘鲁红十字会准备了 100 000 个运尸袋，这一消息更是加剧了民众的恐慌，秘鲁首都的旅游收入和房价大幅下降。美国政府最终紧急派遣一个由科学家和外交官组成的小组前往秘鲁，以缓解当地的紧张气氛。然而，1981 年并没有发生秘鲁大地震，甚至连小地震都没有，这一消息当时占据了各大报纸的头版。

加利福尼亚州，帕克费尔德

利马地震的错误预警给当地人在心理上和经济上都造成了巨大打击，却仍然没有阻挡地震学家继续追求“圣杯”的脚步。在利马的例子中，预测者布雷迪有点像一只孤独的狼，他独自做出预测，还有很多其他预测则得到了美国地质调查局和地震学界坚定不移的支持，只是这些支持多半徒劳无功。

加利福尼亚的帕克菲尔德是被研究得最多的地震带之一，位于弗雷斯诺市和贝克斯菲尔德市之间的圣安地列斯断层上。帕克菲尔德似乎每隔 22 年就会发生一次地震：1857 年发生一次，1881 年发生一次，接下来就是 1901 年、1922 年、1934 年和 1966 年。一篇由美国地质调查局资助的论文设想这个趋势会继续下去，便大胆地预测在 1983~1993 年间会发生一次地震，最可能发生地震的时间是 1988 年。事实上，下一次大地震直到 2004 年才降临帕克菲尔德，与此前预测的时间相去甚远。

帕克菲尔德的地震预测不仅是一个错误，似乎它还强化了一种关于地震的错误观点。这一观点认为地震的发生是有规律的，若某个地区一个时期内没有经历过地震，那迟早要“中招”。板块断层带受到挤压，导致地震发生，压力得到释放，随后又会持续累加直到再次得到释放，就像喷着废水的间歇喷泉，释放压力，恢复平静，如此往复循环。

但断层系统其实相当复杂，像加利福尼亚这样的地区，处于多个断层上，每一个断层都有各自的分区和附属断层区。当板块某处发生地震

时，该板块受到的压力得到释放，但压力却转移到相邻的断层上或是相同板块的其他区域。另外，除非发生地震，否则很难直接测量出板块所受的压力。

也就是说，如果预测旧金山每 35 年会发生一次大地震，这并不意味着这些地震发生的间隔时间是均等的（类似 1900 年、1935 年、1970 年）。更加保险的做法是，假设旧金山地区每年发生地震的概率为 1/35，不论前一次地震是何时发生的，这一概率都不会随着时间推移而发生太大的变化。

加利福尼亚州，莫哈韦沙漠

布雷迪和帕克菲尔德预测失败的案例，似乎让人们在一段时期内不愿再付出努力对地震进行预测，但在 21 世纪初，预测热情又强势回归，新发明的数据导向法在地震预测中颇为流行。

其中一个方法是由弗拉迪米尔·克里斯–布鲁克提出的。这位出生在苏联的地球物理学者年近 90 岁，现任加利福尼亚大学洛杉矶分校教授。克里斯–布鲁克作了很多努力推动地震成因理论的发展。1986 年克里斯–布鲁克第一次招来恶名。那是在雷克雅未克的峰会上，里根总统和米哈伊尔·戈尔巴乔夫会面时，克里斯–布鲁克递交给里根一厚摞报告，报告中预测未来 5 年美国将会发生一次大地震，这一事件随后被人们解读为 1989 年撼动旧金山的洛马·普雷塔大地震。

2004 年，克里斯–布鲁克和他的团队宣称已经在地震预测领域取得重大突破，通过辨识某一区域小地震发生的模式，可以预测到大地震。克里斯–布鲁克识别这些模式的方法复杂繁琐，晦涩难懂，用 8 个方程式代表已发生的地震，在可观测到的时空间隔里，几个方程式又彼此结合应用。该团队称，用这个方法，他们成功地预测到了 2003 年加州圣西蒙地震和日本北海道地震。

这只是克里斯–布鲁克和他的团队的一家之言，圣西蒙和北海道的地震预测究竟是不是在地震前就公布的，还不得而知。查阅学术大全数据库中与报纸相关的部分会发现，2003 年的报纸对这两次大地震只字未提。评价一种预测方法是否成功，关键是要把“预测过去”和“预测未来”区分开，前者可以说是一个逆喻，很显然不能算作成功预测。

而 2004 年 1 月，克里斯–布鲁克则毫不掩饰地向公众公布了自己的又一项预测：在接下来的 9 个月内，加州南部的莫哈韦沙漠地区会发生一场震级至少为 6.4 级的地震。这个预测得到了广泛关注，《发现》杂志、《洛杉矶时报》以及其他很多主流出版物都为克里斯–布鲁克设置了专栏。施瓦辛格州长办公室也打来了电话，还召开了一次紧急会议。

甚至连一向持怀疑态度的美国地质调查局当时都欣然对克里斯–布鲁克表示出一定的赞许，在其官网上肯定克里斯–布鲁克团队的工作是采取合理的方法所进行的地震预测研究。

但是，莫哈韦沙漠地区那一年并没有发生大地震，甚至将近 10 年过去了，该地区也没有发生过一次地震。克里斯–布鲁克团队仍在从事地震预测工作，对加州、意大利和日本的地震进行预测，但鲜有准确的时候。2010 年一项调查发现，该团队提前公布的地震预测中只有 3 次准确，其他 23 次皆以失败收场。

印度尼西亚，苏门答腊岛

地震预测中的另一种错误是，认为某一地区绝对不可能发生某一级别的地震，但后来却发生了。戴维·鲍曼是克里斯–布鲁克以前的学生，现为加州州立大学富勒顿分校地质学系主任。2004 年苏门答腊岛发生 9.2 级大地震，引发的海啸致使 23 万人丧生，此后，戴维·鲍曼将更多的精力投入地震预测。与克里斯–布鲁克的预测方法相同，鲍曼采取的也是数据导向法，利用中等级别的地震预测高级别的地震。然而，鲍曼的方法更加简捷巧妙，更加雄心勃勃，提出了“加速力矩释放”理论，试图测定某一个断层体系中不同点所受压力的大小。另外，和克里斯–布鲁克的方法不同的是，这一系统能预测断层中任意一点发生地震的可能性，不仅能预测到哪里会发生地震，还能预测到哪里不会发生地震。

起初，鲍曼和他的团队确实取得了一些成功：2005 年 3 月，一场震级为 8.6 级的大余震袭击苏门答腊岛，震中就在鲍曼测出的高风险区域。然而，2006 年鲍曼发表的一篇论文表明，在这一个断层中的另一个区域发生地震的可能性较低，这片区域位于印度洋，邻近印度尼西亚的明古鲁省。然而就在一年后的 2007 年 9 月，一系列地震就发生在这片区域，最高震级达到 8.5 级。幸运的是，这些地震离近海很远，伤亡较少，但却推翻了鲍曼的理论。

2007 年“加速力矩释放”模型失败，鲍曼并没有像其他预测者那样责怪运气不好（其实，他的模型也考虑过明古鲁市附近可能会发生地震，只是认为可能性不大），而是重新检查了自己的模型，认定用这种方法预测地震是完全错误的，于是决定弃用。

“我是一个失败的预测者，”2010 年鲍曼告诉我说：“我做了一件大胆却愚蠢的事。我进行了一个可试验的预测，我们应该这么做，可一旦出错，就要自食其果。”

鲍曼一直有一个想法，通过识别地震的成因——地质断层上累积的压力——对地震做出预测。实际上，鲍曼想要了解的是，在整个断层体

系中，压力是如何变化和蔓延的。他的这一观点是受混沌理论的启发得来的。

混沌理论是一个可以制服的恶魔，气象预报员就成功地做到了这一点，至少在一定程度上做到了这一点。气象预报员对大气的理论认识要比地震学家对地壳的了解深刻得多，他们或多或少地知道天气是如何变化的，这种认识甚至深入到分子层面，但地震学家却没有这样的优势。

鲍曼评论道："对气象系统的研究来说，想要了解大气变化是一件很简单的事，只要抬头看看天就好。而我们观察的是岩石，大多数研究对象都在地下 15 000 米的地方，我们不能期望钻个洞探下去，说实话，那只是科幻电影中才有的情节。这是一个根本性的问题，我们没有办法直接测量地壳的压力。"

因为缺乏这样的理论认识，地震学家只能采取纯统计方法预测地震。你可以像鲍曼一样，在自己的模型中创设一个名为"压力"的数据变量。由于无法直接对其进行测量，"压力"这一变量只能表达为过去发生的地震的一个数学函数。而鲍曼认为这样的纯统计方法根本不可能奏效，"数据集里满是噪声干扰，光有统计学意义的数据并不足以验证假说"。

在地震预测和部分的经济学、政治学领域中，有很多干扰数据，理论也不发达。这些领域的知识预测常会按照两步法进行：第一步，将干扰噪声误认为有用信号；第二步，噪声信息流毒于学术期刊、博客媒体、新闻报道，并发出错误的警示，破坏科学，阻碍我们理解这些系统的真实情况。

过度拟合模型：将噪声误认为信号

在统计学中，将噪声误认为信号的行为被称为过度拟合。

假设你是一个小偷小摸的惯犯，而我是你的老板，我要求你想出一个撬密码锁的办法，就是中学学校里常见的那种锁——也许我们打算去偷学生的午饭钱。我期待的撬锁法是，能让我们无论何时何地都可以很有把握地撬开锁。我给了你 3 把锁进行练习，红色的、黑色的和蓝色的。

用这几把锁反复试验了几天后，你回来告诉我，你发现了一个绝对正确的办法：如果锁是红色的，密码组合就是 27–12–31；如果锁是黑色的，密码组合就是 44–14–19；如果锁是蓝色的，密码组合就是 10–3–32。

但我告诉你，你根本就没有完成任务。的确，你想出了开这 3 把锁的办法，但你并没有作更多的努力提升撬锁理论，这个理论可以指导我

们撬开随意一把陌生的锁。我一直非常想知道有没有一种办法可以撬开所有的锁，或者锁本身有什么结构缺陷可被我们利用，或是有哪些破解密码的技巧，比如，某一类数字更常被用作密码，等等。但是，你却给了我一个过于具体的办法，妄图解决一般问题。这就是过度拟合，它是导致预测错误的原因。

拟合的说法源自统计模型和过去的观测结果相吻合的程度，当过于粗略地拟合时（我们称之为不充分拟合），这时我们会错过本可以捕捉到的信号。当拟合太过紧密时（我们称之为过度拟合），这意味着你在对数据中的噪声进行拟合，而不是挖掘数据的深层结构。在实践中，过度拟合的错误更为常见。

为了弄清楚这种错误的成因，我们暂且赋予自己一种在现实生活中几乎不可能拥有的优势——我们清楚地知道真实数据原本的面目。在图 5–4 中，我画了一条平缓的抛物线，顶点落在中间位置，末端逐渐减弱。这条抛物线用来代表你希望了解的所有现实数据。

然而，我们还是不能直接观察到数据的深层关系，这种关系由一系列独立的数据点体现出来，我们只能从这些点推断出这种关系模型。另外，这些数据点还会受特殊环境的影响（有信号，也有噪声）。在图 5–4 中，我画了 100 个数据点，分别用圆形和三角形标记，貌似这样就足以从噪声中捕捉到信号。尽管这些数据中存在一定的随机性，但显然它们仍遵循着这条抛物线。

然而，当我们的数据相对有限时（现实情况常常如此），情况将会怎样呢？那时，我们就更有可能陷入过度拟合的麻烦中。在图 5–5A 中，我将 100 个数据点缩减至 25 个，这时，你会如何连接这些点呢？

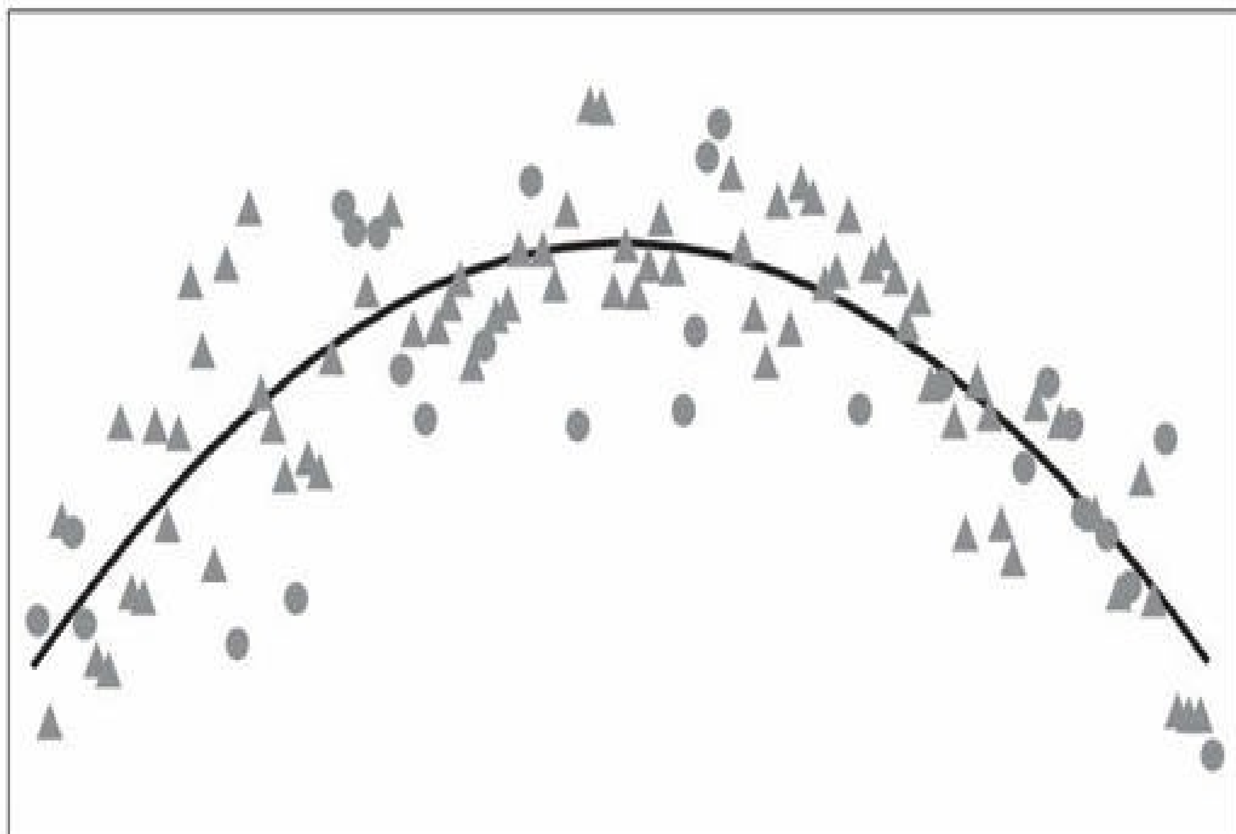


图 **5–4** 数据的真实分布情况

当然，如果你知道真实数据应该呈现的关系模型，就会很自然地将它们连接成一条抛物线。确实，用二次方程式这样的数学表达式可以很好地重建真实的关系模型（如图 5–5B）。

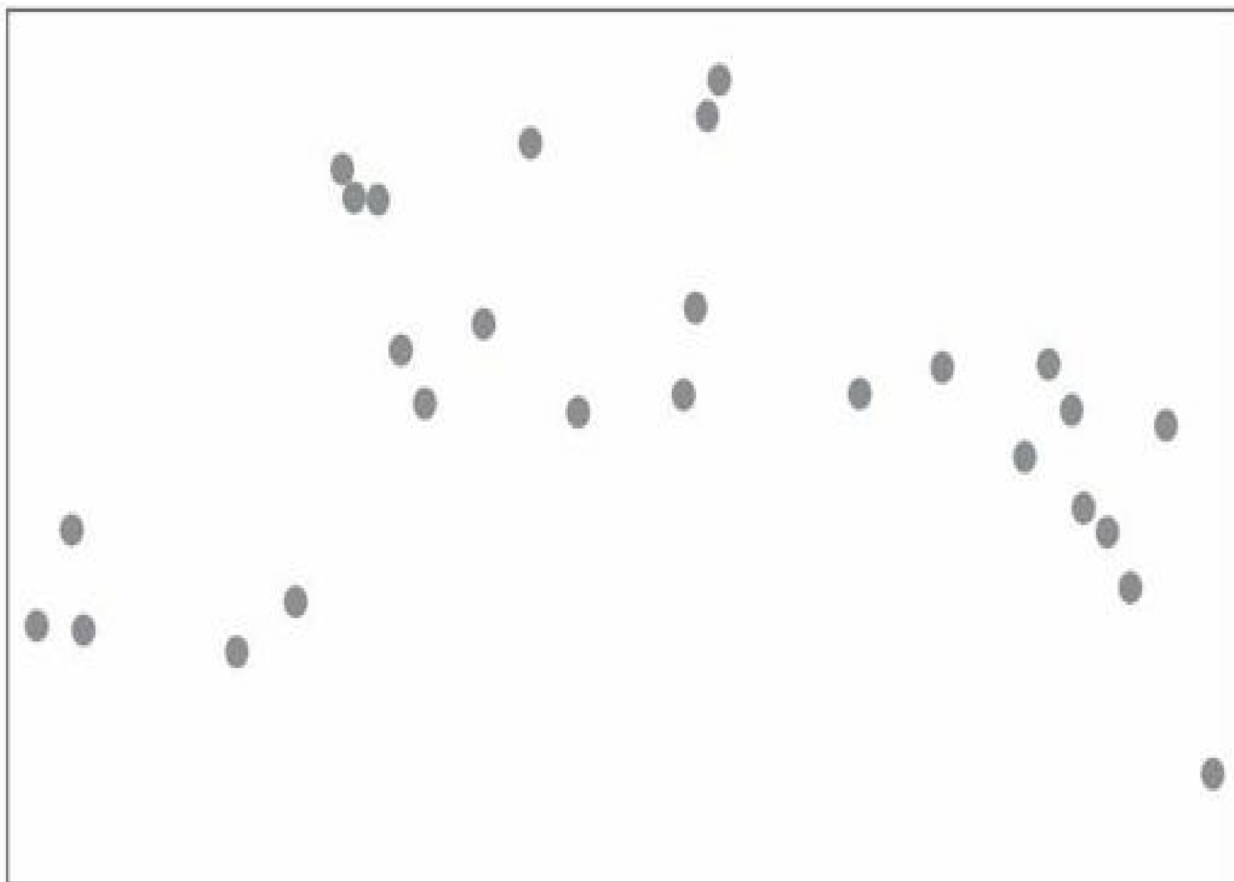


图 **5–5A** 有限数据样本

然而，在无法获知数据的理想模型时，我们有时就会变得很贪婪，图 5–5C 代表的就是这样一个例子，一个过度拟合的模型。在图中，我们设计了一个复杂的函数，可以追踪每一个边缘数据点，用这个函数将这些点连接起来，曲线的上下波动陡然增强。这使得我们离真实的关系模型越来越远，也会使预测更离谱。

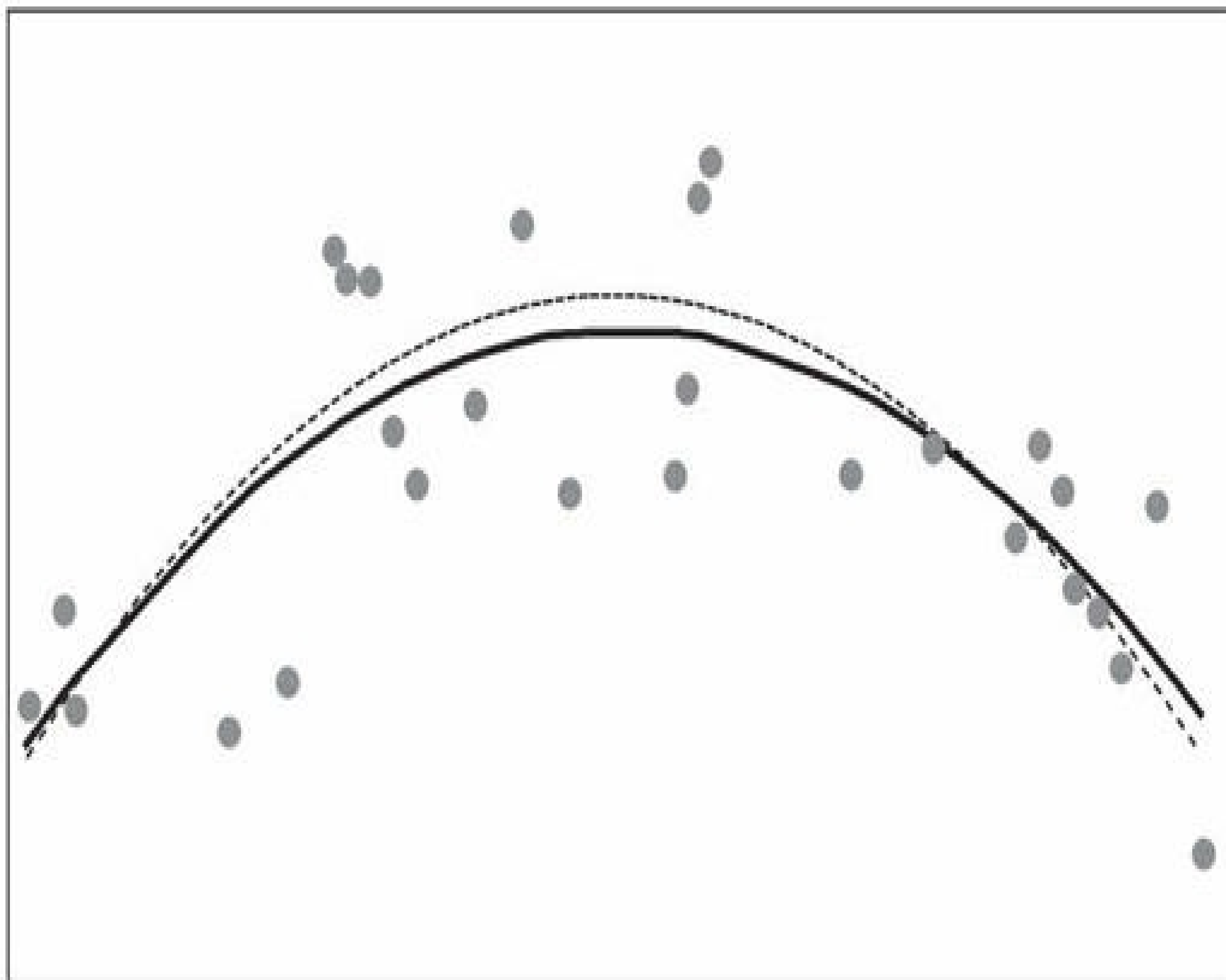


图 **5–5B** 适度拟合模型

这个错误貌似很容易避免，如果我们无所不知，对数据的深层结构总是了如指掌的话，这个错误也确实可以轻松避免。然而，几乎在所有的现实工作中，我们都必须利用归纳法，从已知的证据中对其结构进行推断。当数据有限又充满噪声时，当我们对基本关系的理解很浅显时，就更有可能对一种模型做出过度拟合，在地震预测中，这两种情况可谓司空见惯。

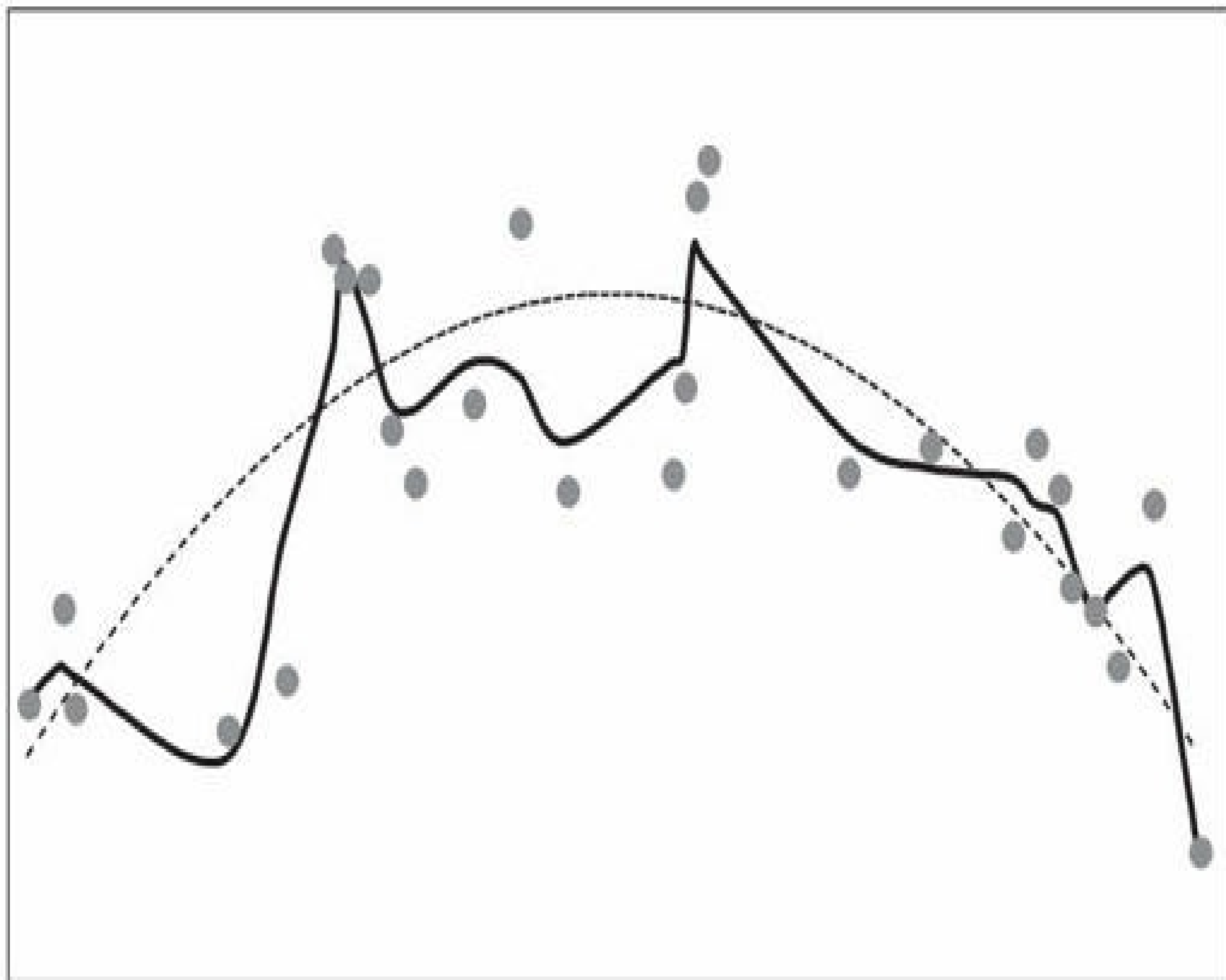


图 **5–5C** 过度拟合模型

如果我们既不知道也不在乎数据关系的真相，就有很多理由可以解释我们为什么倾向于过度拟合的模型。其中一条就是，在预测者最常用的统计测试中，过度拟合模型更受青睐。有一项常用的测试是用来测量我们的模型可以解释多少数据的变化。根据这一测试，过度拟合的模型（图 5–5C）解释了 85%的数据变化，而适度拟合模型只解释了 56%的数据变化。但是实质上，过度拟合模型是在混淆视听，将噪声误当作信号混入模型中。事实上，在解释真实世界时，它的表现更糟。

这种解释似乎让情况一目了然，但很多预测者完全无视这个问题。研究者拥有很多统计方法，可这么多的方法却没有让他们增加一点科学态度，减少一点幻想，而是像充满幻想的孩子在天空中寻找动物形状的云一样。数学家约翰·冯·诺伊曼谈到这个问题时曾说："我用 4 个参数就能拟合出一头大象，用 5 个参数就可以让这头大象甩动它的鼻子。"

过度拟合代表了双重霉运：过度拟合的模型表面上来看比较好，但其实际性能却很糟糕。因为后一种因素，若被用在真实的预测活动中，

过度拟合模型最终会让预测者付出沉重的代价。而因为前一种因素——其表面效果不错，而且自称可以做出非常准确并且新闻价值很高的预测，比其他应用技术都先进，所以，这类模型更吸引人，更容易在学术期刊上得到推介，也更容易被推销给客户，从而将其他可靠的模型排挤出市场。但是，如果这个模型是用噪声拟合的，就很有可能会阻碍科学发展。

你也许猜到了，克里斯–布鲁克的地震模型就是很严重的过度拟合模型。这一模型将一组超级复杂的方程式应用在噪声数据中，也因此付出了代价——预测正确率只有 3/23。戴维·鲍曼意识到他的模型也有类似的问题，便果断终止使用。

应该说明的是，这些错误通常都是真实的。借用另外一本书的题目《随机漫步的傻瓜》[1]来说就是，这些错误常影响我们，让我们更容易被“随机性愚弄”。我们也可能会越来越迷恋这种模型的特质，甚至可能会创造出一个貌似很有说服力的理论来证明这些错误的合理性，自己却没有意识到这一点，这不仅是愚弄同事、朋友，也是在愚弄自己。迈克尔·贝雅克曾就此写了大量的文章，他对这种进退维谷的情况作了如下解释：“在科学中，我们用怀疑主义平衡好奇心。”过度拟合的例子说明是好奇心左右了我们。

2011 年日本大地震引发的思考

将噪声误认为信号的倾向，有时会给现实世界带来极可怕的后果。在日本，尽管地震活动极其频繁，但 2011 年那次灾难性的地震还是让这个国家措手不及。福岛核反应堆是按照可抵御 8.6 级地震的标准设计的，无法承受震级高达 9.1 级的地震。考古资料表明，2011 年由地震引发的约 40 米高的海浪在历史上曾经引发过多次海啸，但这次人们明显忘记了或者根本无视这些惨痛的案例。

9.1 级地震在全世界都很少见，没人能确切地预测到这样的地震会发生在哪个 10 年里，更不用说具体的日期了。而在日本，一些科学家和中央计划员却排除了近期发生地震的可能性。这一点就反映了日本的地震预测模型是过度拟合模型。

在图 5–6A 中，我画出了 2011 年日本福岛地震震中附近发生地震的历史频率。这些数据中的地震级别逐渐加大，却没有达到 3 月 11 日的 9.1 级。通过看图你会发现，数据几乎遵循着古登堡–里克特法则的那种直线预测模型，然而，在 7.5 级处出现了一个拐点，而且，该地区自从 1964 年发生了一场震级达 8.0 级的地震后，再没有发生过震级更大的地震，于是，这条线似乎开始向下弯曲了。

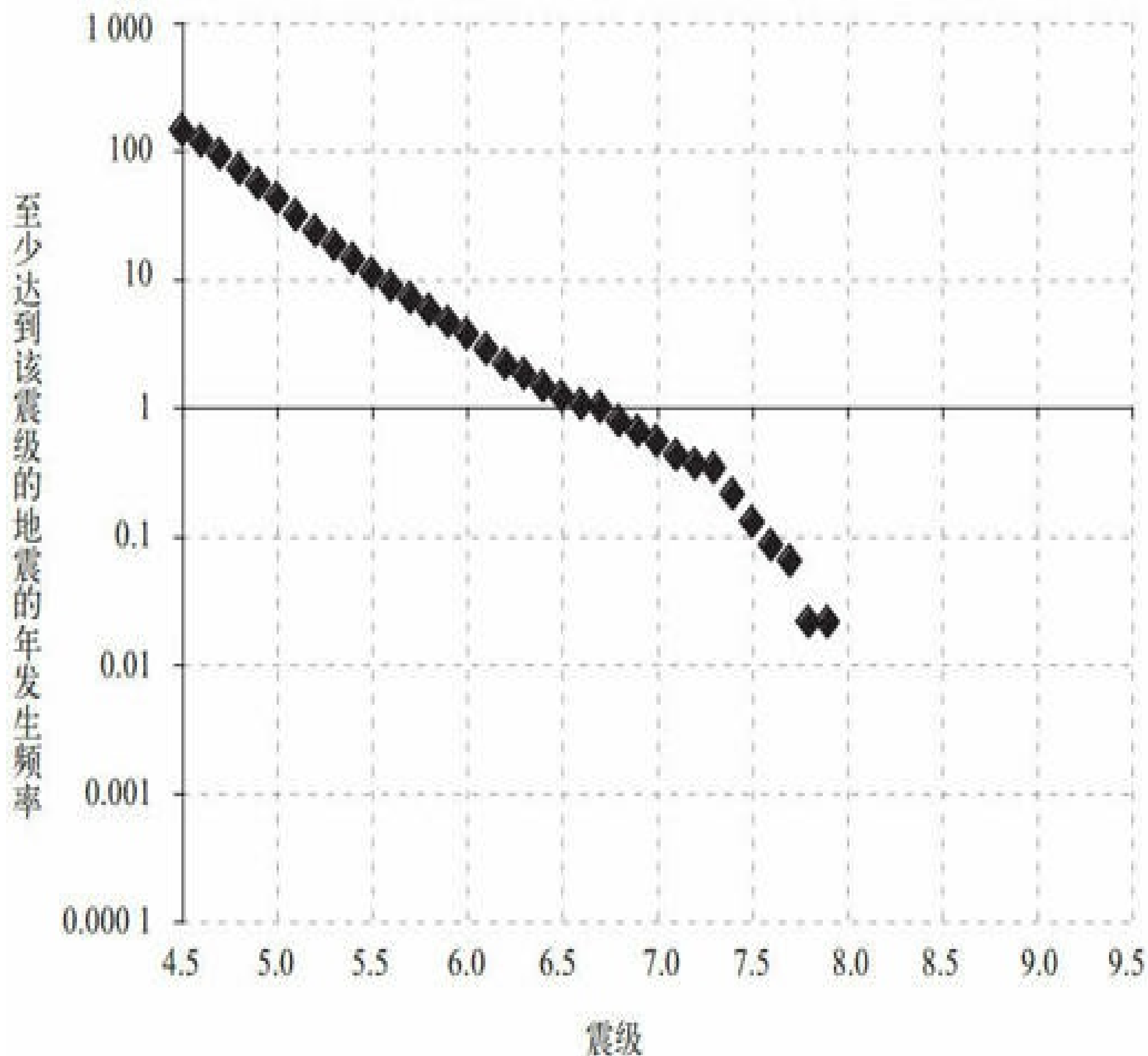


图 **5–6A** 日本东北部地区地震发生的频率图（**1964** 年 **1** 月 **1** 日 **~2011** 年 **3** 月 **10** 日）

究竟该如何连接这些数据点呢？如果严格依据古登堡–里克特法则，就要忽略图像中的拐点，沿直线将数据点连接起来，如图 5–6B 所示。若是按地震学家口中的“特性拟合”法（见图 5–6C），即描述这一地区地震发生的历史频率，那么，就会把那个拐点当成是真实情况，十分肯定这一地区发生 7.6 级以上地震的可能性不大。

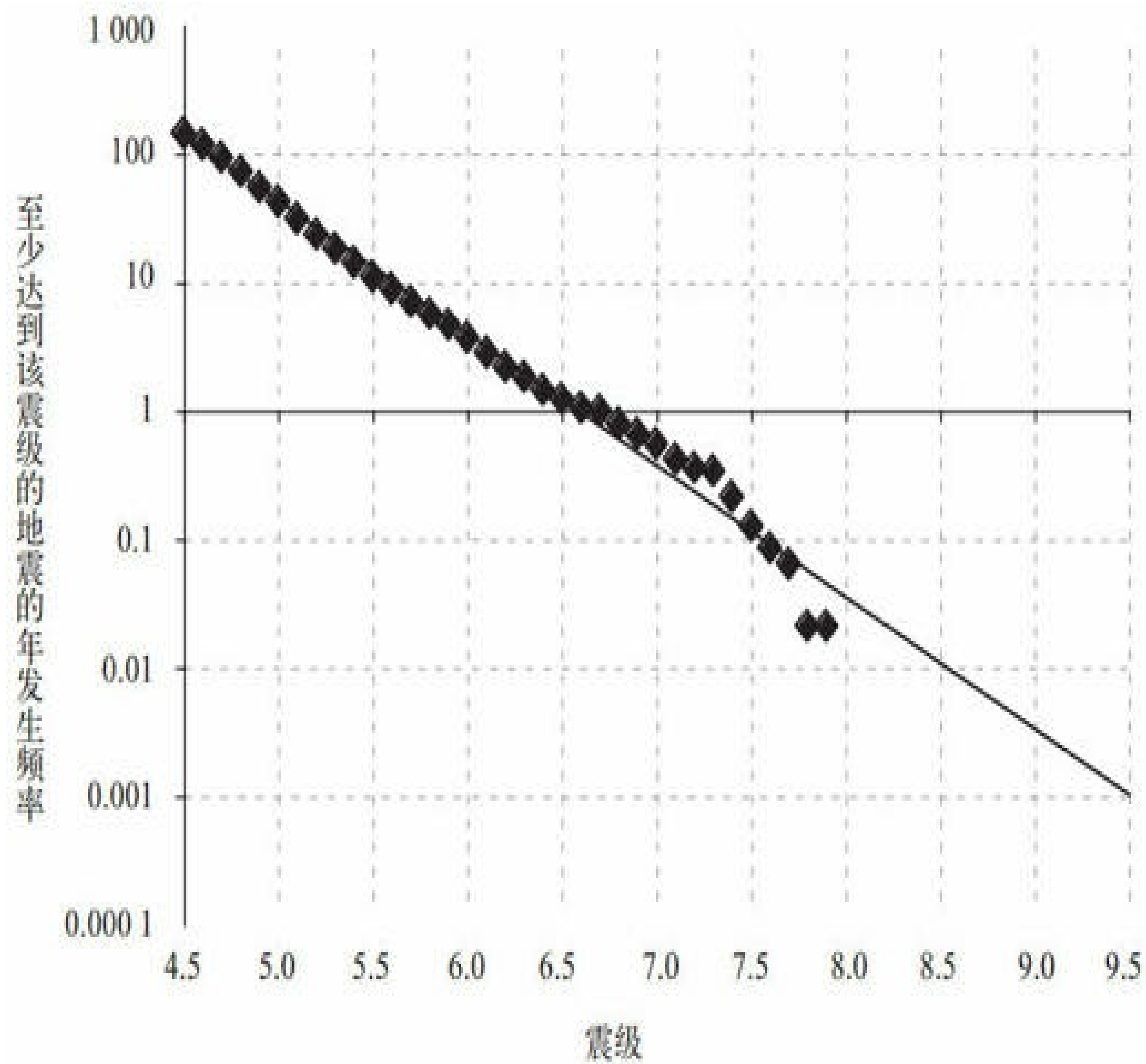


图 **5–6B** 古登堡 – 里克特拟合模型下日本东北部地区地震频率图

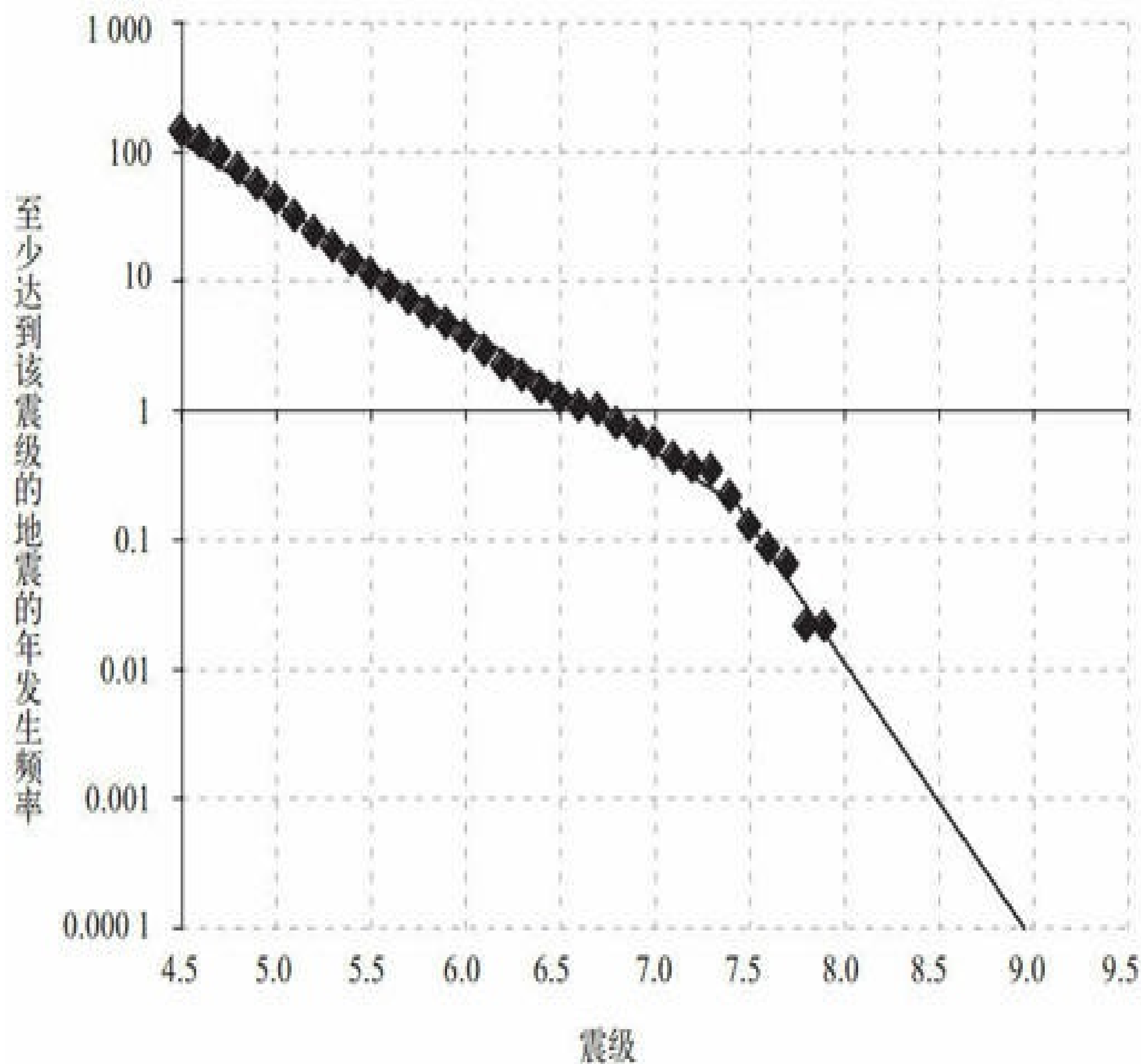


图 5–6C 特性拟合模型下日本东北部地区地震频率图

看似无害的决定，却会导致大相径庭的结果，从日本福岛的例子来看，不同的选择会关系到是否认为这一地区会发生 9.1 级大地震。特性拟合模型暗示 9.1 级地震需要约 13 000 年才可能发生一次，这样看来，这一地区几乎就不可能发生这样大级别的地震。另外，古登堡–里克特法则却预测，9.1 级地震平均每 300 年才可以预见一次，确实不常见，但也不是绝无可能，风险还是存在的，而像日本这样富有的国家，是能够为此作好准备的。

最近，特性拟合模型和日本东北部地区的地震记录拟合得更加紧密了。但是正如我们所知，完全吻合未必就是一件好事，很有可能会变成过度拟合模型，而且，在匹配真实关系时，这种模型会表现得更糟。

在这种情况下，过度拟合的模型会严重低估这一地区发生灾难性大地震的可能性。特性拟合的问题就在于它依靠的是十分微弱的信号。之前提到，日本福岛地震发生前的 45 年内，日本东北部地区从未发生过 8 级及以上的地震。然而，这些都还只是以稀有事件为开端的事例：古登堡–里克特法则预测日本东北部地区约 30 年才会发生一次稀有事件，而 30 年一遇的事件拖到 45 年的时候发生也没什么稀奇，就算没有发生，也不奇怪。就好比一个击球率达 0.300 的击球手，某天状态不好，输赢比达到了 5∶0，这也不足为奇。另外，日本东北部地区发生过好几次中等级别达到 7.0 级的地震，当世界其他地区出现类似情况时，常预示着会有更大级别的地震发生，那又有什么理由认为日本东北部地区会是个特例呢？

实际上，日本以及其他地区的地震学家为此做出了合理解释。比如，他们表明该区域的海底构造特殊、年代久远、温度相对较低且密度较大，能够阻止这样的大地震形成。另一些地震学家则观察到 2004 年以前，这种类型的海底从来没有发生过 9 级地震。

这类结论有点儿像认定了某个来自宾夕法尼亚的家伙不可能会中彩票的头等奖，因为过去的 3 周无人获此大奖。9 级地震就像彩票头奖，中奖人数少，时间间隔长。实际上，2004 年之前，全世界有记载的 9 级地震总共才发生 3 次。9 级地震究竟在何种情形下才会发生，关于这个问题，没有足够的数据来支撑高度具体的结论。日本不是这一模型失败的首例，苏门答腊岛也遇到过相似的问题。有一段时间，苏门答腊岛发生了一系列震级达到 7 级的地震，随后却并没有发生更大级别的地震。但在 2004 年 12 月，一场震级达 9.2 级的特大地震袭击了苏门答腊岛。

古登堡–里克特法则并不能预测地震的具体时间，无论是苏门答腊岛地震还是日本福岛地震都是这样，但是这一法则考虑到了发生地震的概率。目前，许多更精细的地震预测活动都失败了，古登堡–里克特法则的表现都还不错。

地震震级的上限是多少？

近几年发生的特大地震引发了地震学家的思考，地震震级的上限是什么？图 5–2B 呈现了自 1964 年来全世界（包括苏门答腊和日本东北部地区）所有地震发生的频率，你会发现，所有的数据点几乎连成了一条直线。而 10 年前，你可能更关注图形中的拐点（如图 5–6A 日本东北部地区图中的那个拐点），并由此得出结论，特大地震并没有古登堡–里克特法则预测的那么多。但近几年的地震记录表明，特大地震发生得

越来越频繁。

但由于过于罕见，9 级地震的实发频率需要几百年才能知道，至于 9.5 级以上震级的地震，则需要更长的时间。休告诉我，断层区的地质分布会对地震规模产生基本的约束力。如果地球最大、最长的断层带同时破裂，也就是从南美洲南端的火地岛一路向北到阿拉斯加阿留申群岛的所有断层带同时震动，此时地震级别将达到 10 级。但我们很难确切地知道这个震级地震的爆发程度。即使有 1 000 多年可靠的地震记录，我们也无法确定 10 级地震到底是什么样，发生频率又是怎样。也许，地震本身就存在局限性。

地震本身是一个复杂的过程。复杂性理论是由已故物理学家佩·巴克与他人共同创立的，尽管人们经常将这一理论和混沌理论混为一谈，但二者之间是有差别的。复杂理论认为，当一个简单的事物和其他事物互相作用时，就会表现得神秘怪异。

巴克最爱举沙堆的例子。如果一粒沙（有什么比一粒沙更简单呢？）落入一个沙堆中，有可能会发生 3 种情况。依据沙堆的形状和大小，这粒沙可能会停留在落下的位置；或者它会沿着沙堆的斜坡缓缓地流到沙堆底部；还可能出现另外一种情况：如果沙堆太陡，一粒沙就可能撼动整堆沙子，使沙堆崩塌。复杂的系统似乎都有这样的特性，会有很长一段明显的停滞期，而这种停滞状态总是伴随着突发性和灾难性的失败。这些过程也许真不是随机的，而是具有不可简化的复杂性。因此，一旦这种复杂性超越了某种水平，我们就不可能对这些过程做出预测。

被审判的预测科学

当你将镜头推远，从远处看，复杂的过程也会呈现出一种秩序之美。在本书中，我在使用噪声和信号这两个术语时，并不是十分严谨。这两个术语源自电气工程学，那些工程师能识别不同类型的噪声，这些噪声都很随机，但仍遵循不同的基本概率分布。若去听听真正的白噪声，会发现这种声音是由于频率均匀分布的信号上叠加了一些突发声音产生的，是咝咝的声音，也有些像粗糙的摩擦声。与复杂系统相关联的噪声称作布朗噪声（也称红噪声），这种类型的噪声更纾缓，听起来像极了流水的声音。

另外，噪声之美还在于，地壳构造力不仅形成了断层带，还锻造了惊险的山峰、肥沃的山谷、秀美的海岸线。尽管这些断层带潜藏着诱发地震灾难的危险，但人们很可能还会在这里世世代代地居住下去，永不离开。

拉奎拉地震极具讽刺意味的一点是，2011 年 7 位科学家和公职人员因过失杀人罪被告上法庭，拉奎拉市的检察官指控他们未尽全力告知民众地震群后存在发生“大地震”的危险。

显然这是一场荒唐的审判，但这些科学家当初的预测难道就真的没有什么可以改进的地方了吗？或许是有的。已有充分证据表明，地震群发生之后，发生大地震的风险在逐渐增加，可能暂时比基准概率高出100~500 倍。虽然如此，这种风险还是超低的（绝大多数地震群不会引发大地震），但一味暗示民众生活一切正常，大家可以坐下来喝杯小酒，是明显不对的。

本书的观点是，预测者的首要职责就是忠于真理。从广义上看，政治因素会阻碍这种忠诚的行为。地震学界现在仍对利马和帕克菲尔德两地地震预测的失败心有余悸，也害怕与朱利安尼这类人竞争。这就使得他们的工作动机复杂起来，无法专注于自身使命。于是，不负责任的糟糕预测大行其道，准确的预测倒毫无立足之地。

休认为人们永远都无法得到地震预测的“圣杯”，这个观点也许是对的。尽管只有个别地震学家的工作认真负责，但我们还是得到了该学科的汇总资料，可以对其进行评估，这些资料构成了地震可预测性的上千种假说。跟踪记录表明，大多数假说都失败了，而地震预测中的“魔法子弹”也并非什么灵丹妙药。

然而，科学的跟踪记录总的来说是成绩斐然的，这份记录也是一个清晰的信号。用同一种方法反复尝试而不做改变，不可能产生不同的结果。这样做或许没有什么错，但是，科学常常会创造很多“不可预测”的突破。

地震学家在短期地震预测方面取得了一些进步，这可能和拉奎拉地震有关。继古登堡–里克特法则之后，该领域中被广为接受的学术发现就是大地震之后总会发生数次余震。我和一些地震学家交流过，包括加州大学戴维斯分校约翰·朗德尔和南加利福尼亚大学的汤姆·乔丹，这些人都更加关注短期预测，并越来越倾向于一种观点，即应当将这些短期预测的结果清晰完整地传达给公众。

比如，汤姆·乔丹的研究表明，有时余震会按照一个可预测的地质方向，沿着断层带移动。若它们朝着人口密集区移动，即使震级不高，也还是会对生命财产造成很大的威胁。例如，2011 年新西兰基督城发生了 5.8 级地震，185 人丧生，而这场地震却是 2010 年 9 月该国一个偏远地区发生的 7.0 级地震的余震。说到余震，显然会有很多信号，自然会吸引更多的关注。

我们相信，科技终会不断发展。美国宇航局和加州大学戴维斯分校

的朗德尔最近联起手来，采用 GPS（全球定位系统）这样的遥感技术来测量断层所受压力。这些技术颇具发展前景，虽然目前这些技术还不成熟，但地震学家会掌握越来越多的数据，这样他们就能进一步了解地震发生的根本原因，这是很有可能实现的。

这些方法终将推动科学的发展。过去的 1 000 年，地震预测几乎没有一次是成功的，气象预报的准确度也是直到最近 40 年才有所改善，之前也常常预测得不准。又或许，随着对复杂性理论的了解逐步深入（复杂性理论本身是一个崭新的科学分支），我们会得到一个愈加笃定的结论，那就是地震的确不可预测。

无论怎样，一开始很可能都会遭遇很多失败的预测。但错误的记忆会逐渐淡去，那时，信号又会重新在地平线上闪耀，我们仍会如饥似渴地追求成功的预测，即便那只是海市蜃楼也无妨。

[1]《随机漫步的傻瓜》，纳西姆 · 塔勒布，中信出版社于 2012 年出版。——编者注

第六章
经济预测：经济学家为什么没有预测到 2008 年经济危机？

政治选举预测总是被报道存在误差幅度，这让我们看到了其中的不确定性。大多数时候，经济预测都是以单一的数字呈现出来，比如，经济政策下月将会创造 150 000 个就业机会，2013 年的国内生产总值（GDP）将会增长 3%，原油价格将提高到每桶 120 美元，诸如此类。

这些数字会让人觉得这些经济预测惊人地准确。于是，各类报刊常会在报道中对经济预测出现的微小误差表现得大惊小怪：

“失业率意外剧增至 9.2%，引发市场阵痛。”

《丹佛邮报》，2011 年 7 月 9 日

如果你看过那篇字号超小的文章，就会发现这个“意外”的结果是，失业率的确达到了 9.2%，而不是经济学家此前预测的 9.1%。如果 0.1% 的误差就足以上头条新闻，看来这些预测通常必须做到非常可信才行。

然而，经济预测最多不过是一个生硬的工具，很少能提前几个月预知经济的转折点。实际上，更常见的情形是这些预测甚至连已经开始且正在进行中的经济衰退都无法“预言”。最近的 3 次经济衰退——1990 年、2001 年和 2007 年经济衰退——都是这样的情形，衰退明明已经开始，但大部分经济学家却不乐意承认，直到最后才确认这一点。

对美国经济这样复杂庞大的事物进行预测，是一项极具挑战性的工作。这些预测的实际表现和人们对它们的看法之间存在着巨大的差距。

一些经济预测者不愿意让人们知道这一点。与其他大多数学科的预测者一样，经济预测者将不确定性视作仇敌——威胁其名誉的仇敌。他们没有对不确定性做出准确的预测，而是武断地认为预测模型中的不确定性降低了，但在现实生活中，他们却无法提高预测的质量，这样做也无法改进真实世界中的预测。一旦洪水来袭，我们将毫无准备。

不可忽视预测中的不确定性

1997 年 4 月，北达科他州大福克斯市的北雷德河河水泛滥，河水漫过了市镇的堤坝，涌进市内长达 3.2 公里。尽管没有人员伤亡，但该城市 5 万居民几乎都得撤离，清理工作的费用高达几十亿美元，城市中

75%的房屋遭到了破坏或损毁。

与飓风和地震不同，大福克斯市的洪水本来是一场可以避免的灾难：这座城市的防洪石堤本该用沙袋加固，泛滥的洪水也本该被引流到人口稀少的地区——引流到农田里去，而非学校、教堂或住宅区。

大福克斯的居民其实在灾难发生的几个月前就意识到洪水的威胁了。那年冬天，平原地区的降雪格外多，美国国家气象局在积雪开始融化时就预计到会有径流，他们预言雷德河的河水最高会涨到 14.9 米深，接近历史最高纪录。

只是还有一个小问题，大福克斯的那些堤坝当初是按照可抵御深达 15.5 米的洪水的标准建设的，而这次的预测（14.9 米）哪怕出现很小的失误，都会酿成大祸。

结果，雷德河的水位涨到了 16.5 米深。不管怎么说，美国国家气象局的预测都称不上完美，但是，提前两个月做出的预测，出现约 1.6 米的误差也是合情合理的，这样的预测在历史上比比皆是。结合美国国家气象局以往对洪水预测的准确率来看，预测中的误差幅度约为±2.7 米，这便意味着堤坝被冲毁的概率为 35%。

问题是，美国国家气象局明显是刻意对公众隐瞒了预测中的不确定性，只是一味地强调水深将达 14.9 米这一预测结果。负责此次洪灾预测工作的预测者后来告诉那些研究人员，如果自己的观点中传达出任何含有不确定性的信息，他们都担心公众会因此对预测失去信心。

实则不然，如果预测者这样做了，公众反而能准备得更充分——尽可能地加固堤坝、将洪水引流以抵御这次洪灾。然而，若是让民众自行决定，许多人就会深信根本没什么可担忧的（几乎没人购买洪水保险）。气象专家毫无保留地预测此次洪峰为 14.9 米，似乎是在暗示人们，此次洪水的最大深度就是 14.9 米，分毫不差，15.5 米高的堤坝足以保护公众的安全。有些居民甚至将 14.9 米的预测值解读为洪峰可能达到的最大值。

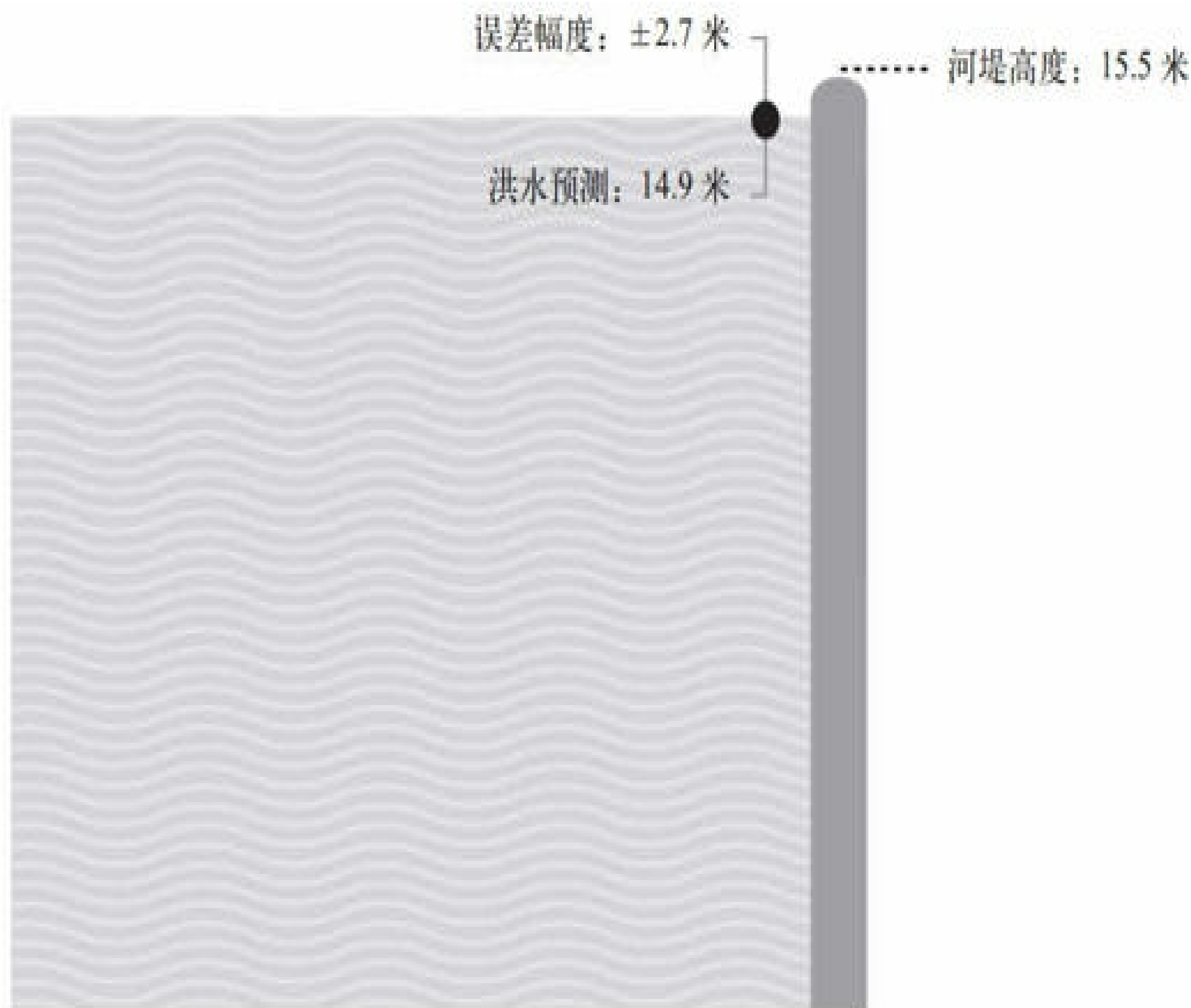


图 **6–1** 洪水预测的误差幅度

有这样一个被引用了无数次的笑话：一位统计学家趟过一条平均水位不足 1 米深的河时，竟被淹死了。此次洪灾，美国国家气象局的预测模型给出的水位平均值是 14.9 米，但水位只要稍高一点，就会淹没整个城镇。

自此以后，美国国家气象局开始逐渐意识到向公众准确、诚实地传达不确定性的重要性。然而，其他领域的预测者几乎都没有秉持这种态度，在预测经济运行状况时尤其如此。

经济学家都是理性的吗？

现在，我们来回忆一下 2007 年 11 月都发生了什么事情。那是经济大衰退正式来临的前一个月，房地产市场已经出现明显的危机迹象：房产止赎率上涨了 1 倍，全美的抵押贷款机构全都濒临破产。信贷市场也出现了不祥的预兆。

费城联邦储备银行推出的一项关于“专业预测人员情况调查”（以下简称“调查”）的季度民调显示，很多经济学家认为经济衰退不大可能发生，相反，他们预期经济仍会缓慢增长，只不过略低于 2008 年 2.4%的平均增长率。他们还认为，发生严重经济衰退的概率几乎为零，殊不知当时一场严峻的经济衰退即将上演。

这项“调查”的独特之处在于，它要求受访的经济学家明确地指出经济发展的方向及可能产生的结果。正如我在本书中一直强调的那样，概率性思考是科学预测必不可少的组成部分。假设在你掷两个骰子之前，我让你说出所有可能出现的情况，正确的答案绝不会是一个单一的数字，而是列举出所有可能出现的数字以及各个数字出现的概率，如图 6–2 所示。假设从长远来看，每个数字出现的概率都与你的预计一致，尽管掷出数字 7 的可能性最大，但也未见得就能始终掷出这个数字，掷出数字 2 或 12 的概率也是同样大的。

“调查”要求受访的经济学家对 GDP 和其他经济变量进行预测时，也要考虑到概率问题，比如，估算 GDP 的可能值是 2%~3%还是 3%~4%。图 6–3 显示的是 2007 年 11 月，经济学家对 GDP 做出的预测：

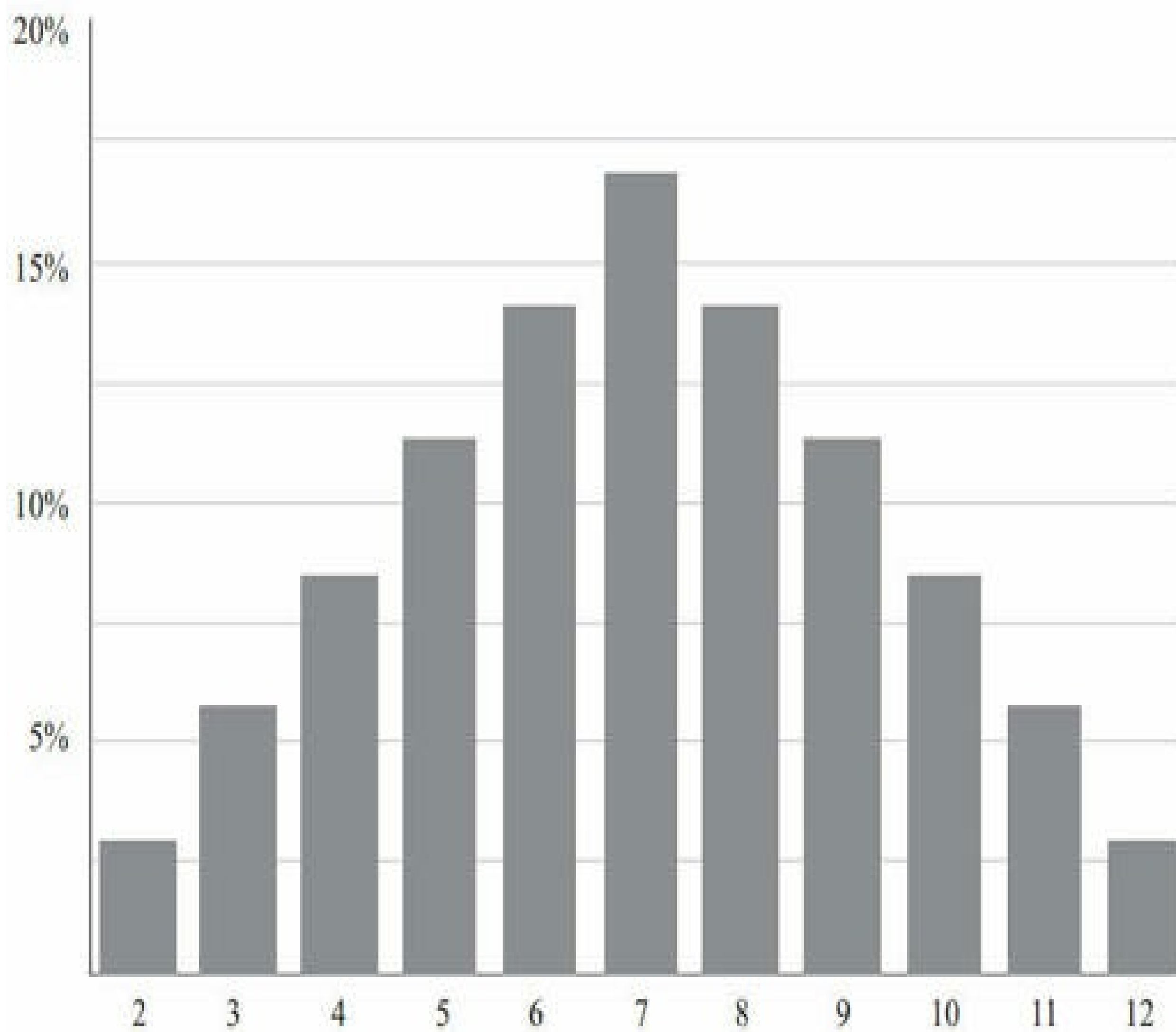


图 **6–2** 预测结果的概率分布：投掷两个骰子可能出现的各种结果

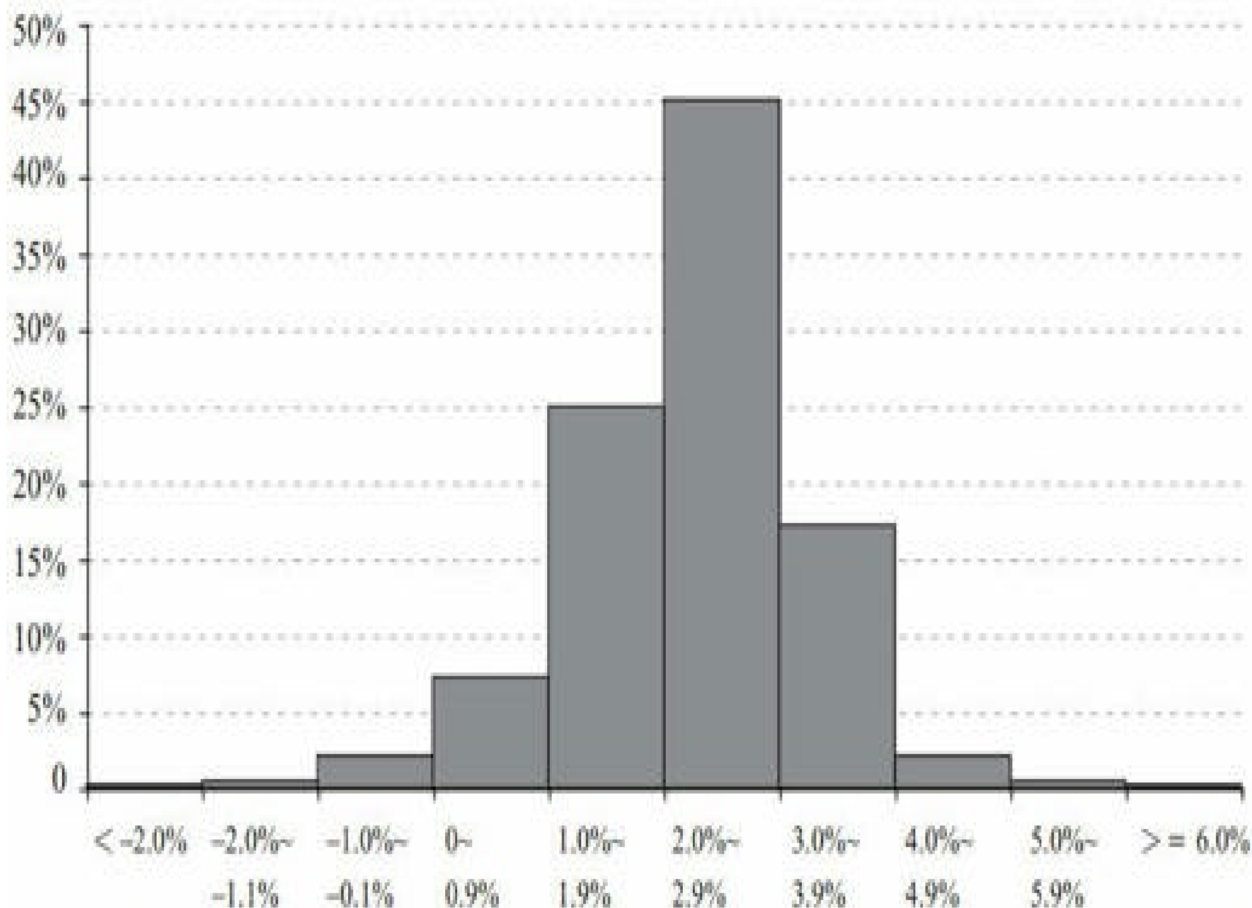


图 **6–3** 预测结果的概率分布：美国 **GDP** 的真实增长水平（**2008** 年）

我在前面提到过，“调查”中受访的经济学家认为，2008 年 GDP 最终会实现约 2.4%的增长，略低于长期增势。这是一个非常糟糕的预测：2008 年金融危机爆发后，美国的 GDP 非但未增，实际上还缩减了 3.3%。更可笑的是，这些经济学家对其糟糕的预测竟然还满怀信心，他们认为，整个 2008 年经济出现缩水情况的概率不会超过 3%，GDP 缩减 2%的概率也只有 2‰，实际上，当年的 GDP 确实缩减了 2%。

的确，长期以来，很多经济学家对于自己预测经济发展趋势的能力都太过自信。在图 6–4 中，我标出了 1993~2010 年这 18 年间所有“调查”对 GDP 增长率的预测值。图中的竖条代表的是经济学家给出的预测区间中 90%的范围。

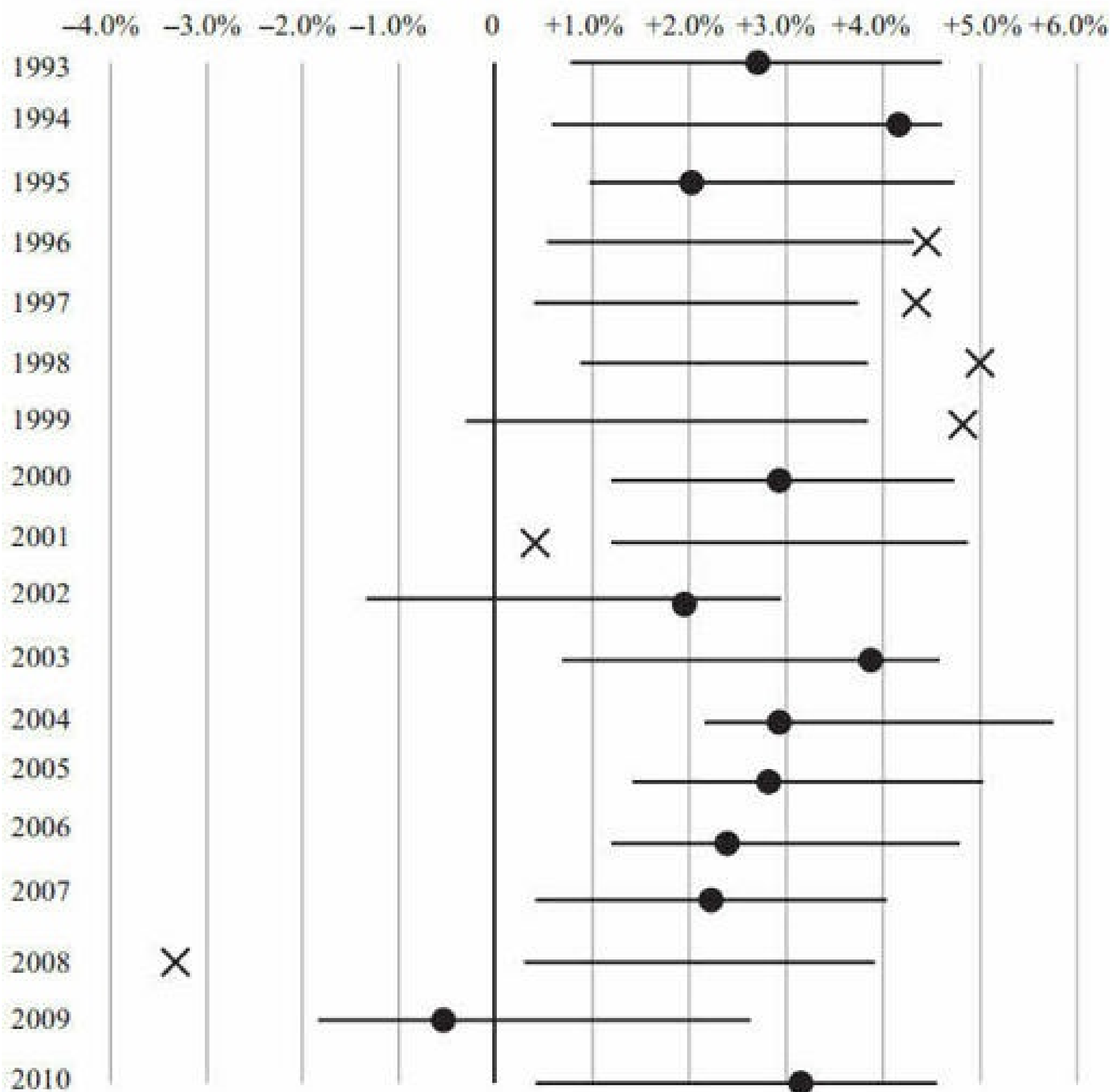


图 **6–4 GDP** 增长率预测：**90%** 的预测区间与实际情况对照表·

一个预测区间包含一系列由预测提供的有可能出现的结果，很像民调中的误差幅度。比如，90%的预测区间应该对现实生活中可能出现的结果实现 90%的覆盖率，而剩下那 10%的无关紧要的情况会被分配到预测区间的边缘位置。如果这些经济学家的预测如其所言，十分准确，那么预测区间里 GDP 增长率的实际数值十有八九是不会下跌的，或者说 18 年里只有两年 GDP 会出现下跌。

但实际上，GDP 的实际数值在 18 年里 6 次跌出了这些经济学家的预测区间，可以说这些预测的错误率足有 1/3。另一项研究将这些数据追溯到 1968 年，也就是这项“调查”的起始年，得到的结果更加糟糕：

GDP 增长率的实际数值有 1/2 都跌出了预测区间。这就不能说是那些经济学家的运气不好了，而是他们从根本上夸大了其预测的可信度。

在现实生活中，当一群经济学家预测 GDP 时，90%的预测区间可以跨越约 6.4%的真实 GDP 增长率（相当于误差幅度为±3.2%），这要基于预测的实际表现，而非经济学家的一面之词。

当你从新闻中得知，明年的 GDP 将会增长 2.5%，这就意味着 GDP 很可能会以 5.7%的高比率轻松增长，当然也有可能下降 0.7%——一次严重的经济衰退。那些经济学家只能做到这种程度了，而且他们在预测方面似乎也没有什么进步。有这样一则笑话，经济学家对近期发生的 6 次经济衰退进行预测，结果正确预测的次数达到 9 次，这则老生常谈的笑话似乎有点儿道理。一项真实的统计数据表明，20 世纪 90 年代，在全世界 60 次经济衰退中，经济学家提前一年预测到的衰退只有两次。

就这一点而言，经济学家并非特例。这样的做法已成为预测行业的行规：那些专家既不擅长诚实地描述预测中的不确定性，对这种做法也不大感兴趣。过度自信的预测体现出的这种特性在其他许多领域也显露出来，包括医学研究、政治科学、金融学以及心理学。不论是靠人的判断进行预测（如菲利普·特罗克等政治科学家的预测案例），还是用统计模型进行预测，似乎都能找到过度自信的影子。

但是经济学家犯了这种错误后，也许无法像其他领域的专家那样总能找到借口。首先，他们的预测不仅过度自信，而且现实意义也不大，比如在预测实际 GDP 增长率时，误差很大，即经济学意义上的回旋余地很大。其次，有组织地预测 GDP 增长率这类变量的活动已经持续好多年了，可以一直追溯到 1946 年的美国利文斯顿调查，这些结果保存完好，可供人们免费参阅。收集人们对我们所作预测的反馈信息是改进预测的一种方式，也许也是一种基本方式。与大多数其他行业的专家相比，经济预测者能得到更多的反馈，但他们却并没有以此矫正过度自信的倾向。

经济学难道不应该成为研究人类理性行为的领域吗？当然，你可能可以理解其他领域的专家，比如人类学家，在进行预测时会表现出偏向性，但绝不希望经济学家这样。

然而，这其实只是这个问题的一部分。那些经济学家相当了解理性，这意味着他们也相当了解我们的动机是如何发挥作用的。如果他们做出带有偏向性的预测，就说明他们没有多强烈的动机去做出好的预测。

相关的两个经济变量未必互为因果

考虑到上述预测者的跟踪记录，我最想找到的经济学家只有一种：他愿意坦承预测工作并非易事，而且坦承自己的预测很容易出错。而我恰好找到了一位——简·哈祖斯，他是全球性投资银行高盛集团的首席经济学家。

近几年来，可以说哈祖斯至少比他的竞争对手更可信。2007 年 11 月，当大多数经济学家仍旧认为不可能出现经济衰退时，哈祖斯写了一份颇具挑衅意味的备忘录，标题为“举债经营带来的损失：抵押贷款违约为何事关重大”，这份备忘录警告人们，数百万名房产拥有者拖欠抵押贷款，引发信贷市场和金融市场的多米诺骨牌效应，造成数万亿美元的损失以及一场可能非常严重的经济衰退。事实证明，备忘录中提到的大部分情况都实实在在地发生了。同时，哈祖斯及其团队还迅速地降低了他们对危机后经济奇迹般复苏的可能性的估值。2009 年 1 月，美国通过了经济刺激方案，一个月后，美国白宫宣称，截止到 2009 年年底，失业率将降至 7.8%，而哈祖斯则预计，失业率会升至 9.5%（十分接近实际数字 9.9%）。

哈祖斯，一个成熟中略带忧伤的德国男人，在入职高盛 8 年后，于 2005 年成为公司首席经济学家，即使对大银行持怀疑态度的人，对他也会心存敬意。保罗·克鲁格曼曾经这样对我说：“（简·哈祖斯）为人友善，我希望高盛总裁劳埃德·布兰克费恩不要对简和他的团队横加干涉。”在谈到对美国经济走势的预测时，哈祖斯总是很谦逊。

我在纽约西街高盛公司明亮的办公室里与哈祖斯会面时，哈祖斯说：“说起这个问题，谁心里都没有底。商业世界的事极难预料，想要弄懂经济这种复杂的有机组织是非常难的。”

在哈祖斯看来，经济预测者面临着三大基本挑战。第一，单纯依靠经济统计数据，很难判断起因和结果。第二，经济始终都在变化，某一经济周期的经济运行状况无法被用来解释未来经济的发展。第三，经济学家以往的预测如此糟糕，那么他们作预测时所参照的数据也好不到哪里去。

美国政府每年公布的数据，与经济指标直接相关的有 4.5 万个，而私人数据提供者要追踪高达 400 万个统计数据。一些经济学家忍不住想要把所有数据都混合在一起，并给一般的数据穿上优质的“外衣”。第二次世界大战之后只出现了 11 次经济衰退的情况，如果一个统计模型试图解释这 11 次衰退带来的后果，就必须从 400 万个数据中选择数据，由此得出的许多相关性都将会带有欺骗性。（这是过度拟合的又一个经典例子——将噪声误认为是信号——在本书前面的内容中，这个问题发生在地震预测者身上。）

想象一下，当你拥有像电话簿一样厚的一摞经济数据时，自己会变得多么富有创造力。比如，美国橄榄球超级杯大赛的冠军队伍曾一度是经济预测中非常有名的“领先指标”。从 1967 年的第一届超级杯大赛到 1997 年的第 31 届超级杯大赛，当来自原美国国家橄榄球联盟（NFL）的队伍赢得比赛时，那么股市就会平均上涨 14%，而如果是来自原美国职业橄榄球联盟的队伍赢得比赛时，那么股市就会平均下跌 10%。

1997 年之前的 31 年里，有 28 年该指标都正确地“预测”了股市的走向。从理论上来看，统计学标准测试可以表明这种相关关系绝非偶然，是偶然的可能性只有 470 万分之一。

然而事实上，这种关系的出现只是巧合。最终，这一指标的表现还是退步了。1998 年，原美国职业橄榄球联盟的丹佛野马队赢得了美国橄榄球超级杯大赛，按照惯例，这应该是一个不好的预兆，然而，这次股市非但没有下跌，反而在网络公司的强势带动下上涨了 28%。2008 年，原美国国家橄榄球联盟的纽约巨人队奋力直追，外接手戴维·泰里大放异彩，打乱了新英格兰爱国者队夺冠大卫轮胎队的计划，然而，新英格兰爱国者队的失利并没能阻止房地产泡沫的破灭，这次房市崩溃导致股市暴跌了 35%。事实上，自 1998 年以来，每次原美国橄榄球联盟的队伍赢得超级杯大赛时，股市非但不会下跌，反而会有约 10%的上涨幅度，与传说中这一指标的预测走势截然相反。

为何一个本来失败概率只有 470 万分之一的指标会一败涂地？出于同样的原因，尽管美国的全国性彩票强力球彩票的中奖概率达到 1.95 亿分之一，但每隔几周还是会有人中奖。对于所有买彩票的人来讲，中奖概率都是相当小的，但数百万张彩票一经出售，总会有人走运中奖。同样的，在世界上数百万统计指标中，有一些恰好与股票价格、GDP、失业率有很大的相关性。如果不是超级杯的冠军，就可能是乌干达的鸡肉产量。这种关系仅仅是一个巧合而已。

尽管那些经济学家可能并不会真的把美国橄榄球超级杯大赛的结果当成领先指标，但他们还是说服自己去相信其他类型的变量——所有与经济意义沾边儿的事物都算——也可以成为提前几个月预测经济衰退或复苏的重要“领先指标”。曾有一家预测公司吹嘘自己对 400 个领先指标了如指掌，远远超过了哈祖斯所说的包含大部分经济事物在内的二三十个领先指标。其他预测者则把一些相对晦涩的指标，如预订量和提货量的比率，高价卖给那些半导体公司。有这么多经济变量可以选择，其中有些变量一定可以将噪声拟合进历史数据。

想要找到能识别信号的变量就更难了，一个经济周期中作为领先指标的变量经常会成为下一个经济周期的滞后指标。2003 年《公司》杂

志中的一篇文章提到的 7 个所谓的领先指标，都曾成功地预测了 1990 年和 2001 年的经济衰退。但其中只有两个指标——房价和临时雇工——能够使人们感知到 2007 年开始的经济衰退。其他指标，如商业贷款，直到 2008 年，即经济衰退开始一年以后，才开始显现出下跌的趋势。

即使人人都看好的“领先指标”也存在问题。“领先指标”由经济咨询商局发布的 10 个经济指标合成。在经济衰退几个月前，“领先指标”就已经普遍地降低了，但它给出的基本上都是假警报，包括 1984 年最臭名昭著的那次，当时这个指数连降了 3 个月，预示着经济衰退，然而经济非但没有衰退，反而以 6%的增长率快速发展。有些研究甚至声称，“领先指标”在实际生活中，根本不具备任何预测作用。

哈祖斯对我说：“真正具有预测性的变量几乎不存在，想要弄明白哪些是因果关系，哪些属于相关性，是很困难的。”

大多数人都听说过这句格言：“相关性并不是指因果关系。”两个变量之间存在统计学关系，并不代表两者互为因果。比如，冰淇淋的销量和森林火灾是两个相关变量，因为两者都会在炎热的夏季出现，但两者之间没有因果关系：当你买了一盒哈根达斯冰淇淋时，并没有点燃蒙大拿的灌木丛。

然而，这一概念表达起来容易，应用起来却十分困难，在理解经济领域的因果关系时尤其如此。比如，哈祖斯注意到失业率一直被视为滞后指标，有时也确实如此。经济衰退之后，许多企业在对经济的前景恢复信心之前，不大可能雇用新员工，因此失业人员要经过很长一段时间才能重返工作岗位。但是，失业率也可以成为衡量消费需求的领先指标，原因是失业人员对新产品和服务的购买力很低。经济衰退期间，经济发展可能陷入这样的恶性循环：在消费需求上升之前，企业不会雇用新员工，失业的消费者就买不起企业的产品，这样又会导致消费需求不足。

消费者信心是另一个捉摸不定的变量，总会变来变去。有时，消费者是最先发现经济预警信号的群体之一，而有时他们又是最后觉察到经济复苏的群体之一，因为一直以来，经济衰退实际上早就结束了，但公众却认为经济仍然处于衰退期。于是，对于消费者信心是领先指标还是滞后指标这个问题，经济学家仍存有争议，判断标准也许要视经济运行所处的周期决定。另外，由于消费者的信心会对消费行为产生影响，在经济预期和实际经济情况之间，可能会有各种各样的反馈环（反作用）。

或许经济预测和经济政策之间的反馈环问题更严重。比如，如果预

测到一场经济衰退即将发生，那么美国政府和美联储可能就会采取行动降低风险，或者减少损失。这样就会产生一个问题，那就是像哈祖斯这样的预测者既要预测政治决策，又要预测经济决策，而对美国这种国会拥有的支持率仅为10%的国家，这无疑是一个挑战。

然而，问题会进一步加剧。正如 1976 年诺贝尔经济学奖获得者罗伯特·卢卡斯指出的那样，经济模型以历史数据作为依据，这些数据一部分来自方针决策，而这些决策在当时是适用的。因此，只知道现在的政策制定者想要做什么是不够的，还需要对尼克松政府执政期间的财政和货币政策有所了解。由伦敦政治经济学院教授提出的“古德哈特定律”认为，政策制定者一旦锁定一个特定变量，这个变量就会逐渐失去其作为经济指标的价值。比如，如果美国政府人为地上调房价，房价会上涨，但房价就不再是衡量整体经济是否健康的指标了。

从逻辑的角度来看，这有点像观察者效应（常被误认为海森堡的不确定性原理）：一旦我们开始测量某物，它的行为就会发生改变。大多数统计模型都是依托这一概念建立的，输入和输出分别是自变量和因变量，彼此之间泾渭分明，但在经济领域中，两者却混在一起、乱作一团。

变化莫测的经济

即使经济学家可以解决这些问题，他们仍然需要应对动态的目标。美国和全球的经济一直在发展，不同经济变量之间的关系会随着时间的推移发生改变。

比如，历史上 GDP 增长率和工作岗位增加率两者之间一直保持着一种合理的密切联系，经济学家称之为“奥肯定律”（或奥肯法则）。1947~1999 年的经济长期繁荣期内，工作岗位增加率通常是 GDP 增长率的 1/2，所以，如果 GDP 的增长率为 4%，那么工作岗位的增长率就为 2%。

这一关系至今仍旧存在——GDP 增长得越多，工作岗位也增长得越多，但它们之间的动态关系似乎有所改变。前几次经济衰退过后，工作岗位的新增数量寥寥无几，比经济长期繁荣期的新增岗位数量少很多。比如，2009 年经济刺激方案出台后，根据奥肯定律，GDP 的增长速度已足够创造约 200 万个工作岗位，而实际上，这一时期工作岗位的数量不但没有增加，反而减少了 350 万个。

这种变化意味着什么，经济学家对此总是争论不休。包括哥伦比亚大学的杰夫瑞·萨克斯在内的经济学家做出的最消极的解释是，GDP 增长和工作岗位增加之间的关系反映出美国经济许多严重的结构问题，比

如，来自别国的竞争不断加强，服务业和制造业之间的不平衡状态，人口老龄化问题，中产阶层萎缩，还有越来越高的国家债务。在奥肯定律下，美国进入了一个全新的非正常状态，如果不进行根本上的改变，这些问题会更加严重。萨克斯告诉我，“我们低估了全球变化对美国变化的影响。国际上的工作岗位都转移到了中国，新兴市场确实动摇了美国经济。”

更大的问题是，21 世纪初的经济波动是否更能代表长期的经济发展状况——也许经济长期繁荣一直是表面现象。1947~1999 年的 50 多年时间里，经济处于衰退期的时间只占整个时间段的 15%，但 1900~1945 年这 45 年里，衰退期所占时间的比例为 36%，是繁荣期所占时间的两倍还多。

尽管大多数经济学家认为，我们在稳定经济周期方面已经取得了一些进步，但我们或许只是很幸运地躲过了更多的问题。这一点在 1983~2006 年这段时间尤为明显，衰退期只占整个时间段的 3%，所以这一时期有时也被称作“大缓和”时期，是经济长期繁荣期的一个子集。但是，经济增长依靠的大多是激增的政府和消费者债务，还有各种资产价格泡沫。再先进的经济也没有神授的权利可以使自身按照“大缓和”时期的增长率发展：20 世纪 80 年代，日本经济的年增长率为 5%，而此后的增长率仅为 1%。

2007 年的经济大衰退危害如此之深，让那些预测人员和政策制定者感到措手不及，原因或许也在于此。他们不但未能解释“大萧条”这样的重大事件，有时还会根据“大缓和”时期的经济状况调整他们的预测。事实上，从历史角度来看，“大缓和”时期不过是非主流现象罢了。

负责制定利率的美国联邦公开市场委员会（FOMC），每年依法需要向美国国会公布宏观经济预测至少两次。到 2007 年年末，比起私有企业的预测者，FOMC 在某些方面的预测更准确，比如他们公布的 GDP 增长预测就略胜一筹，这就使得那些私有企业在 2007 年年底之前，先后 4 次降低利率预测值。

然而，在 FOMC 于 2007 年 10 月底召开的一次长时间的会议中，仍未提及“经济衰退”这个词，即使是在讨论经济问题时，也没有提到。FOMC 在措辞上十分谨慎，尽管这样，“经济衰退”还是可能通过使用“经济下降的危险”这类词体现出来。但是，FOMC 不认为会发生经济衰退（他们预测经济走势仍会呈增长状态），而且也没有迹象表明他们愿意直面当时正处在严峻的经济衰退期的现状。

出现这种情况的一部分原因可能是，为了做出准确的预测，FOMC 参考了“大缓和”时期的数据。FOMC 成员对一篇记录了 1986~2006 年经

济预测表现的论文有些过度依赖。只关注近几年的预测表现所带来的问题就是，这些预测包含的经济波动太少，只有发生在 1990~1991 年和 2001 年的两次较为缓和的经济衰退。论文作者提醒人们："用 20 世纪 80 年代中期的数据估计现在的不确定性，就是在含蓄地假设'大缓和'时期之后的平静状态会一直持续到未来。"这是一个十分糟糕的假设。FOMC 正是因为忽视了那些出现严重经济衰退的年份，才得出 2007 年不会出现严重的经济衰退这一结论。

一个预测者永远都不应该忽视数据，在研究经济衰退或是美国总统选举这样的少有事件时尤其如此，因为这类事件可供参考的数据并不多。忽视数据通常表明预测者过于自信，或是在对自己的模型进行过度拟合——不是为了使预测更加准确，而是为了炫耀。

在这个特殊案例中，经济学家的预测准确性并没有明显提高。在图 6–5A 中，我将"调查"中 GDP 增长的预测水平与 1968~1985 年（这些都是 FOMC 注意过但又忽略掉的年份）的实际增长数字作了对比，你会发现，这一时期确实有很多经济波动，如 20 世纪 70 年代中期和 80 年代初由通货膨胀导致的经济衰退。这些结果并没有让预测者完全沮丧，原因在于预测结果和实际结果之间有着密切关联。

如果对 1986~2006 年的 GDP 增长情况做一个同样的对比图（如图 6–5B），情况正好相反。大部分数据点，不论是有关 GDP 的预测值还是实际数值，都紧密地聚集在年增长率为 2%~5%的狭窄区域里，因为这一时期经济波动非常小，预测的平均误差小于之前的那个时期。然而，经济中存在一定程度的可变性，如 1990~1991 年或 2001 年较为缓和的经济衰退，这份预测确实没有很好地抓住这一点——实际上，这份预测结果与实际结果之间几乎没有半点关联。没有迹象表明经济学家对经济的预测越来越准确，相反，由于经济运行平稳，他们的工作暂时变得更加轻松，这就好比檀香山的气象预报员的工作要比布法罗的预测者容易得多。

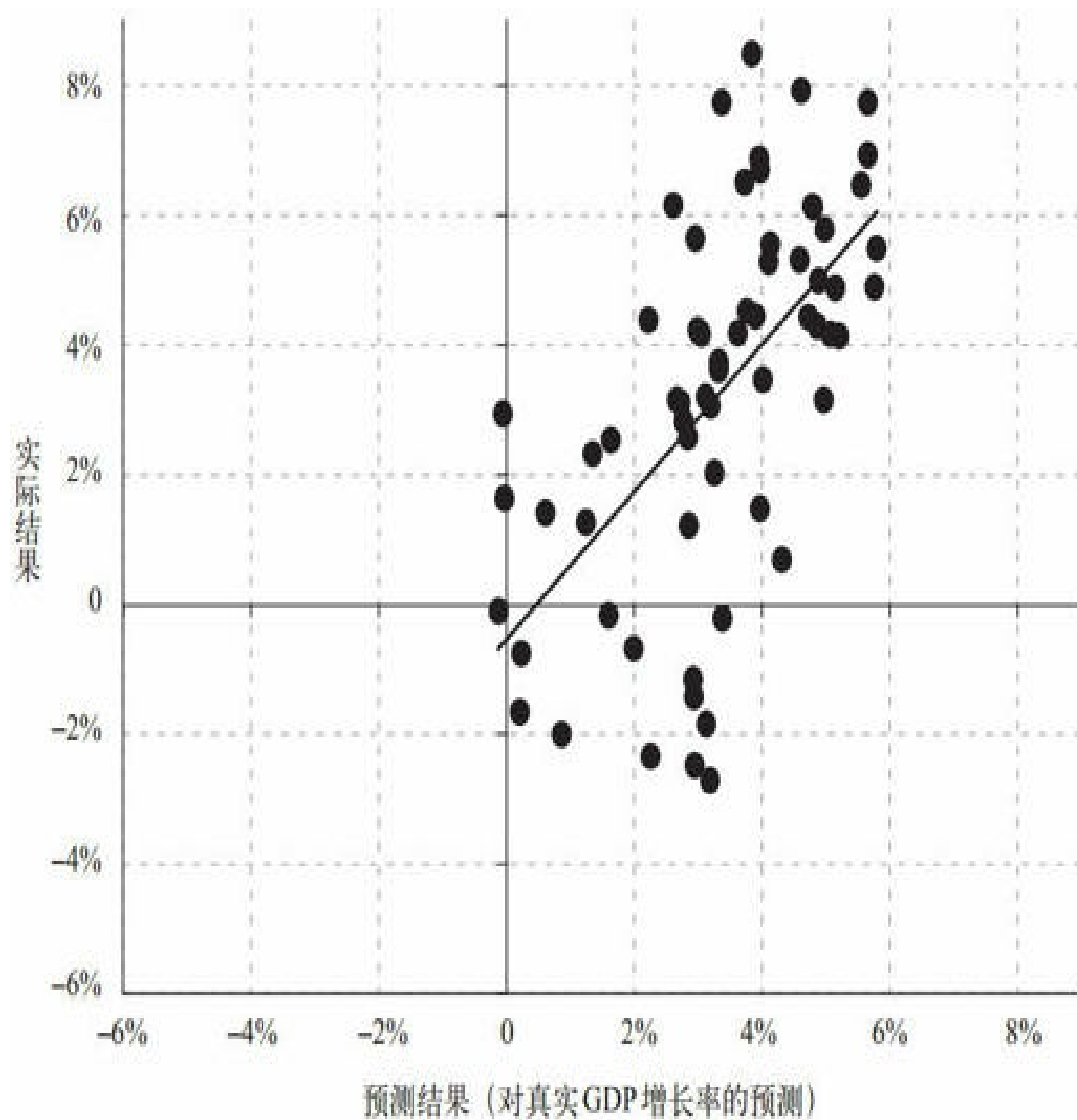


图 **6–5A　1968~1985** 年 **GDP** 预测结果与实际 **GDP** 对比图

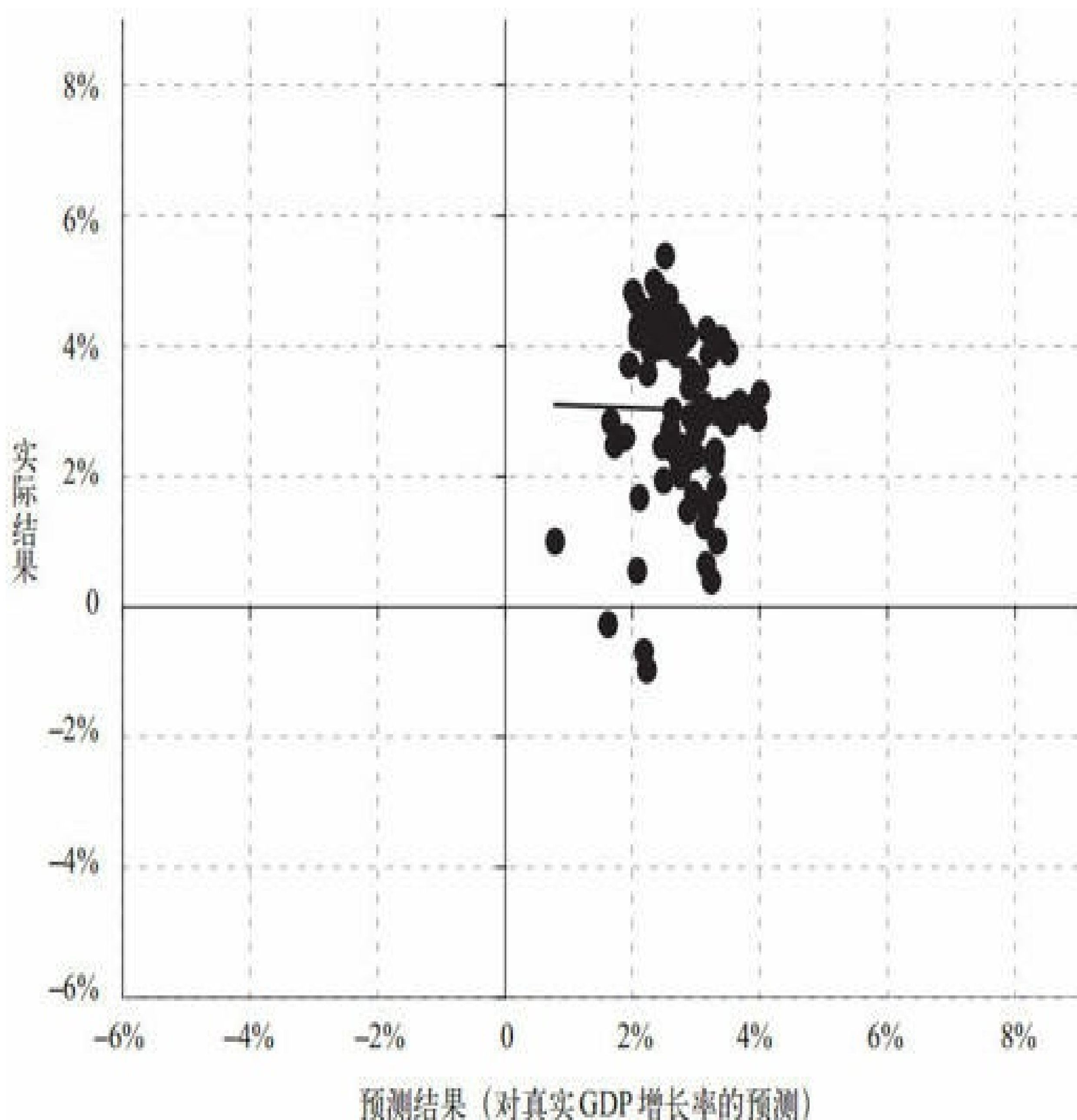


图 **6–5B 1986~2006** 年 **GDP** 预测结果与 **GDP** 实际结果对比图

有时你会听说，抛开数据的另一个原因就是，你正在试图解决的问题发生了根本性的转变。有时，这些论断在某种程度上是说得通的：美国经济在不断发展，也会经历周期性的结构转变（最近，制造业主导的经济正在向服务业主导的经济转变）。经济不是棒球，棒球比赛总是遵循同样的规则，而经济规则不会一成不变。

问题是，你永远都不会知道下一个转变什么时候会出现，也不会知道这个转变会使经济变得更加波动还是更加平稳，更加强大还是更加脆弱。如果一个经济模型是建立在“没什么大变化”的前提条件之上，那么

这个经济模型就毫无价值可言。但是，预测这些转折点的出现实非易事。

经济数据中充满噪声

经济预测者面临的第三大挑战是，他们的原始数据质量不高。我在前文曾经提到，经济预测者在进行预测时，很少公开自己的预测区间，或许是因为这样做会降低公众对他们的专业知识的信任度。哈祖斯说：“经济预测者为什么不公开预测区间呢？因为他们怕出丑。我认为就是这个原因，他们怕出丑。”

然而，不仅经济预测中存在着不确定性，经济变量本身也具有不确定性。大多数经济数据序列都需要修正，统计数据一经公布，修正工作可能长达数月甚至数年。修正工作有时是十分庞杂的。美国政府对2008 年第四季度 GDP 增长的估计就是这样一个不算体面的例子，最初公布 GDP“仅”下降了 3.8%，现在又改为当时经济的降幅将近 9%。如果白宫这些经济学家对经济漏洞的实际大小有所了解，他们就会在 2009 年 1 月强烈推行更大的经济刺激方案，或者他们会意识到问题的严重性，从长计议，而不是试图速战速决。

此类错误相当普遍。1965~2009 年，美国政府对 GDP 的季度增长率的初始估值最终得到修正，平均变动幅度为 1.7%。这只是一个平均变化，而每个季度的变动幅度仍然较大，最初 GDP 的季度增长率估值的误差幅度为±4.3%。这就意味着，即使政府最初报告说经济会出现超常增长，可实际上经济有可能正处于衰退期，反之亦然。比如，1977 年，美国政府公布第 4 季度美国经济增长了 4.2%，但这个数字后来被修正为–0.1%。

所以，我们还是应该同情经济预测者。预测经济发展趋势实属不易，但如果连经济预测从何处着手都不知道的话，那就难上加难了。

经济是一个动态系统，不是一个方程式

经济学家与气象预报员面对的挑战或许有些可比性，他们都面临着以下两个基本问题。

第一，经济与大气一样，是动态系统。不同事物之间互相作用，这些系统都处于永久的运动当中。在气象学中，这个问题非常明显，因为天气系统符合混沌理论。从理论上讲，巴西的蝴蝶扇动一下翅膀，就会在美国得克萨斯州引起一场龙卷风。在经济领域中也存在着大致相同的情况，日本海啸或是长滩港口工人罢工也会影响到某个得克萨斯人是否能找到工作。

第二，气象预报受某些不确定的初始条件的约束。气象预报总是采

用一种或然说的表达方式（“降雨概率为 70%”），之所以采用这种方式，并不是因为天气本身存在随机性，而是因为气象学家认为他们对初始条件的测量不够精确。天气预测模型（由于符合混沌理论）对初始条件的变化极为敏感。同样的，在经济预测中，初始数据的质量通常都非常差。

然而，气象预报仍是本书中真正成功的预测例子之一。由于人与机器的通力合作、更好的数据收集方法和一如既往的辛勤工作，不论是飓风轨迹预测还是日间高温预测，都比一二十年前准确多了。

对于经济预测来说，还不能那么自信。对经济预测的任何幻想，都会被经济学家在金融危机之前所犯的重大错误打破。

如果气象学家也像经济学家一样遇到一些初始条件不明的动态系统问题，他们会用大量的自然科学知识来补救。像龙卷风这类天气现象，其物理属性和化学属性没有那么复杂，但这并不是说龙卷风很容易预测，而是对于龙卷风的形成和消散的原因，气象学家有着深刻的认识。

经济学更偏向于人文科学。尽管经济学家对于控制经济运行的基本体系有着相当正确的理解，然而，经济运行中各种情况的因果关系仍然模糊不清，在经济泡沫期和经济恐慌期时尤其如此，因为此时的经济系统中充斥着人为因素决定的反馈环。

尽管如此，如果对经济学家来说辨别因果关系很难的话，尝试一下总好过放弃。比如，再次仔细考虑一下哈祖斯在 2007 年 11 月 15 日写下的这番话：

> 抵押贷款可能造成的损失所引发的宏观经济风险要远远超出公认的水平……这种宏观经济的后果可能会相当戏剧化。如果负债经营的投资者看到贷款的总损失达到 2 000 亿美元，他们可能就会把借款相应地缩减 2 万亿美元。这是一个很大的打击……显而易见，这样的打击会引发严重的经济衰退，或是长期的经济萎靡。

哈祖斯还写道，因为房地产泡沫的影响，房价已经超出了消费者的购买能力，所以许多消费者用长期的大额贷款买了房，其中很多人已经停止还贷了，而这样很有可能造成极大的贷款损失。负债经营的程度会使问题复杂化，导致信贷市场和金融业出现更大面积的瘫痪。所以说，这样的打击足以引发严重的经济衰退。

这正是金融危机爆发的原因。哈祖斯的预测是正确的，他给出的理由也是合理的，不仅解释了金融危机爆发的起因，还预测出金融危机带来的影响。哈祖斯将这一因果关系链称作“故事”。这是一个与经济有关

的故事，即使这个故事可能是一个数据导向的故事，但也是根植于真实世界的故事。

相反，如果你只把经济当作一系列变量和方程式，而没有看到其深层结构，那就很容易把噪声当成信号，误认为自己（还有那些轻信的投资者）正在做出准确的预测，而实际上你的预测并不准确。看看哈祖斯的竞争对手——美国经济周期研究所（ECRI）是怎么做的。

2011 年 9 月，ECRI 预测到近期可能会出现“双底衰退”（亦称“双谷衰退”），ECRI 认为，“决策者无法阻止这次双底衰退。你认为这就是经济不景气吗？好戏还在后头呢”。在对 ECRI 总经理拉克什曼·阿楚坦的采访中，他认为经济衰退如果还未开始，那么马上就会开始了。对此说法，他们是这样解释的：

> ECRI 的经济衰退预测不是仅仅基于一两个领先指标，而是基于很多特定的领先指标，包括美国的长期领先指标、低迷期的周领先指标还有其他短期领先指标。实际上，那些最可靠、最让人期待的指标现在的集体表现，和完全衰退期的最高峰时的表现一样。

上面的解释包含了很多术语，可唯独缺乏切实的经济内容。ECRI 的故事与数据有关（仿佛是数据本身引起了经济衰退），却与经济无关，他们似乎还引以为荣。在 2004 年出版的一本书中，他们这样建议客户：“你不必为了安全驾驶而去了解汽车引擎如何工作，同样，你也无须为了准确地阅读这些评估标准而去了解经济领域的庞杂繁复。”

进入大数据时代之后，这样的说法越来越常见。有了这么多信息，谁还需要理论呢？但在预测行业中持有这样的态度绝对是错误的，尤其是在数据如此杂乱的经济领域。有了理论支持，或者至少对其根源进行更加深入的研究之后，统计学推理就会变得更具说服力。比如，2011 年 9 月，在欧洲出现债务危机时，对经济持悲观态度是合情合理的，但 ECRI 并不这么认为，它们拥有大量的随机变量，这些变量误把“相关性”当成了“因果关系”。

的确，ECRI 的预测似乎标志着一个经济转折点，但却是朝着积极的方向转变。在 ECRI 发出经济衰退预测的 5 个月后，标准普尔公司的赢利率达到 21%，同时，2011 年第 4 季度 GDP 不但没有陷入衰退，反而实现了 3%的平稳、健康增长。ECRI 采取了缓兵之计，并“澄清”说它们的预测一直延伸到 2012 年，尽管这与它们最初的说法根本就不沾边儿。

经济预测中不可避免地会存在偏见

如果你想进行经济预测，最好的选择就是查看平均预测或群体预测，而不是求助于某个经济学家。我对“调查”的研究显示，群体预测总是比个体预测更准确，在预测 GDP 增长、失业率和通货膨胀这三个方面，群体预测比个体预测的准确率分别高出 20%、10%和 30%。通过研究许多领域的预测结果，人们发现几乎所有的群体预测都优于个体预测。

然而，虽说群体预测优于个体预测这一观念已成为重要的经验性规律，可是当预测与事实有很大的出入时，这一观念有时就会成为蹩脚的借口。群体预测是由个体预测组成的，如果个体预测的质量提高了，群体预测的质量也会提高。另外，在现实生活中，经济群体预测的质量也很差劲儿，所以还有很大的提升空间。

大多数经济学家作预测时，会在一定程度上依赖自己的判断，而不是依据统计模型输出的信息进行预测。考虑到数据是那么杂乱，这种做法或许是有益的。波士顿联储前副总裁斯蒂芬·K·麦克内斯曾经进行过一项研究，他发现根据统计学预测方法对人为的判断进行调整会使预测的准确率提高约 15%。20 世纪七八十年代计算机开始广泛使用时，人们普遍认为统计模型能够“解决”经济预测问题。但是，改进的技术无法掩盖对经济领域理论认识的缺乏，只会让经济学家更加快速、更加煞费苦心地将噪声误认为是信号。看似前景不错的预测模型在某些方面一败涂地，最后惨遭淘汰。在其他领域，比如那一时期的地震预测，也会遭遇这样的状况。

援引某个人为判断也会带来潜在的偏见。人们在进行预测时，会倾向于使预测满足自己的经济动机或政治信仰。人们或许太过自负，即使事实和环境要求他做出改变，他也不愿对自己的预测进行修正。哈祖斯告诉我：“我认为人们绝对有这样的倾向，急切地希望事情能按照自己希望的方式发展下去。”

是否有经济学家更擅长把握这种权衡的度？预测出上一次经济衰退的经济学家是不是也可以预测出下一次经济衰退？这个问题有一个非常有趣的答案。

用于评判预测技能的统计测试，在应用于“调查”时给出了否定的结果。也就是说，从这份调查中，人们似乎找不到证据证明某些经济学家总是比其他同行出色。然而，对另一个小组——“蓝筹经济调查报告”——的研究却总能给出肯定的结果。经济预测中显然少不了运气的成分，暴躁或是顽固的经济学家偶尔也会做出正确预测。但是对“蓝筹

经济调查报告”的多项研究发现，长期来看，某些经济学家似乎确实要比其他同行略胜一筹。

这两份调查有什么不同呢？“专业预测人员情况调查”是匿名操作的：每位经济学家随机分配一个身份标识号码，他们在不同的调查中会一直使用这个号码，但是不会暴露自己的身份和职业。而在“蓝筹预测调查报告”中，每个人的预测都标有其姓名和曾经获得的荣誉。

当预测结果标有自己的名字时，你的动机就会发生改变。比如，如果你在一家名不见经传的小公司工作，你的那些大胆预测就会备受重视，不管这个预测是准确的还是离谱的，你都会受人关注，这是合情合理的。而像高盛这样的大公司，为了能达成一致意见，预测反而会比较保守。

事实上，在“蓝筹经济调查报告”中，已经发现了这个特点：名气越小，作预测时就越不怕冒险。一项研究把这一现象称为“合理偏见”。即使知道这个预测十分冒险，你还是会为“大比分”放手一搏，这样做是可以理解的。反过来，如果你已经声名鹊起，即便自己手头的数据要求你做出大胆的预测，你可能也不愿意做出过于大胆的预测。

这两种对声誉的关注都有可能使你远离做出最诚实、最准确的预测这一目标，甚至还会使群体预测更加糟糕。整体来看，在对 GDP 增长率和失业率的预测上，“专业预测人员情况调查”的匿名参与者略优于参与“蓝筹经济调查”的那些名誉至上的受访者。

克服预测偏见的两种选择

如果说做出失败的预测是合乎情理的，那么这些预测肯定是在一些用户的纵容下做出的。正如一些政治专家以向党派支持者提供虚假消息为生一样，经济领域的投机者也因其特有的经济思想而拥有一定的市场。（有时，经济预测也有明确的政治意图，比如，从历史上看，不论是民主党还是共和党执政，白宫提供的经济预测通常都是最不准确的。）

然而，经济预测的风险要比政治预测高得多。正如罗伯特·卢卡斯所言，经济预测与经济政策之间的界限十分模糊，一个不准确的预测可能会使现实中的经济状况变得更加糟糕。

技术进步还是有望提高经济预测的准确度的。例如，谷歌的搜索流量就可以作为预测失业率这样的经济数据的领先指标。

在位于加利福尼亚芒廷维尤的谷歌总部，公司的首席经济学家哈尔·瓦里安对我说：“在我们看来，是否领取了失业保险中的最初赔偿，这一指标能够很准确地预测失业率，也能准确地预测经济活动。如果你在

某一家公司上班，公司里即将裁员的谣言不胫而走，人们就会纷纷开始搜索‘失业人员救助办公室在哪里’、‘我该如何申请失业救助’等词条，从而更早地预测出申请失业救济的失业人数了。这算是一个领先指标。”

经济学和其他领域的预测历史显示，如果受到人类偏见的影响，技术进步也许就无法提高预测水平，而迄今为止鲜有迹象表明经济预测者已经克服了自身偏见，比方说，那些经历过“经济大衰退”的经济预测者似乎就没有刻意控制个人偏见。看看 2011 年 11 月“调查”对 GDP 增长做出的预测（图 6–6），你就会发现其中显示出的过度自信倾向仍然和我们在 2007 年看到的一样，那些预测者对经济运行的上升和下行态势的预测都估计不足，按照他们预测的历史准确性来看，这次的预测错得有些离谱。

我们永远都无法彻底摆脱偏见，但是如果想减少这些偏见，我们有两个基本选择。一种是“供应性”选择，即为准确的经济预测创造市场。另一种是“需求性”选择，即减少对不准确且过度自信的预测的需求。

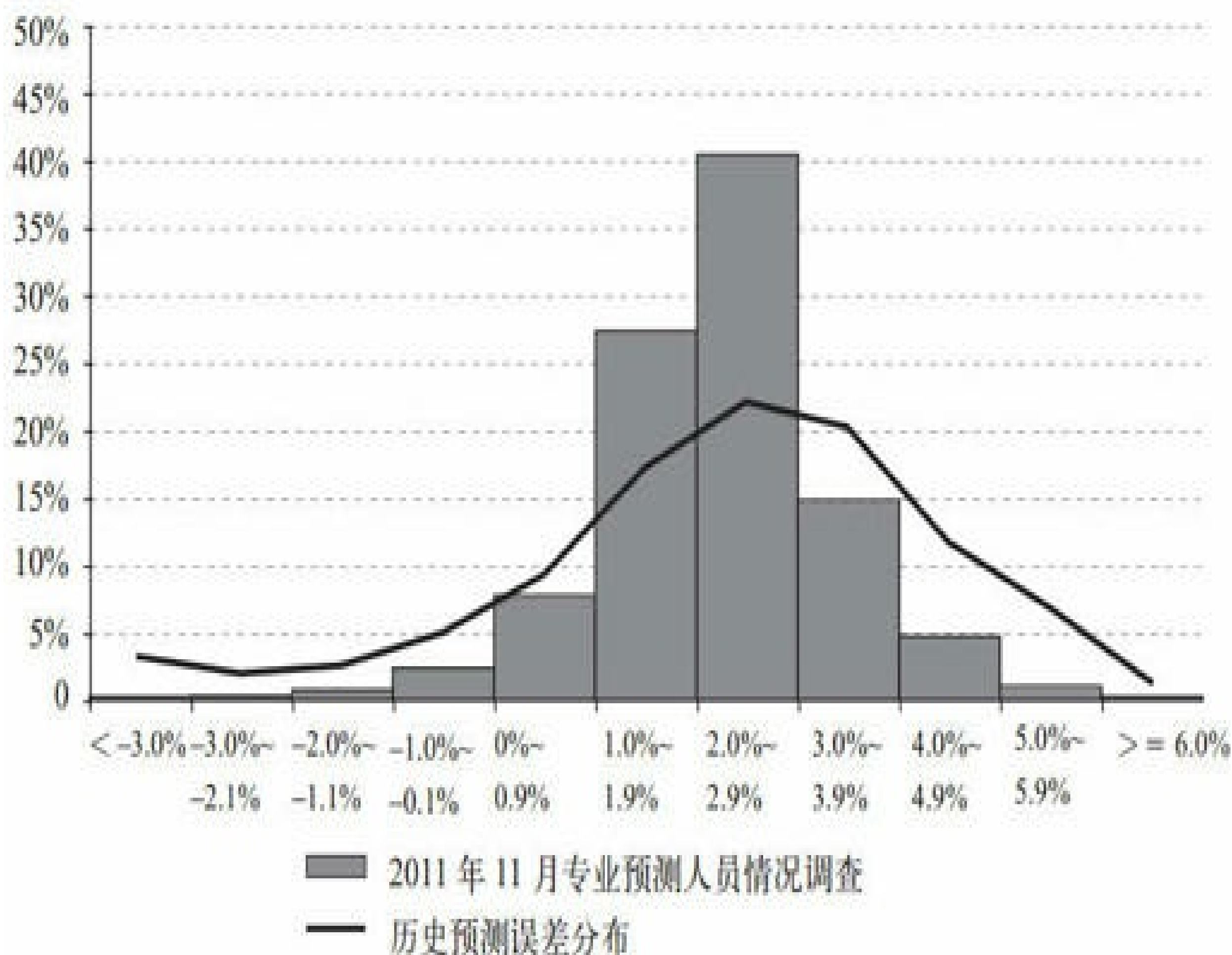


图 **6–6** 预测的概率分布：美国 **GDP** 的真实增长水平（**2012** 年）与历史预测误差对比图

乔治梅森大学的经济学家罗宾·汉森是供应性选择的支持者。我曾经在弗吉尼亚北部的一家摩洛哥餐厅与汉森共进午餐，这是他最喜欢的地方之一。汉森刚过 50 岁，尽管头顶脱发严重，但是看上去很年轻，也有些古怪。他打算死后把自己的头冷冻保存起来。除了供应性选择之外，汉森还支持被他称为"法达奇政府"（一种在一定程度上受投机市场控制的理论上的政府形式）的系统，在该系统中，为政治决策拍板的是预测市场，而不是政治家。汉森显然并不惧怕挑战传统观点，他写过一篇题为"战胜偏见"的博文供读者思考，文化禁忌、思想观念或者错位动机，究竟是哪个因素限制了人们做出最佳决定的能力呢？

服务员端来我们的食物时，汉森对我说："我认为最有趣的问题是，我们为预测付出的努力其实是多么微乎其微，即使对于那些我们认为十分重要的事情也是一样。"

他继续说道："在一个工商管理学硕士学院，你看到的管理者都是大决策者的形象或科学决策者的形象，他们每天忙于电子数据表和统计测试，权衡各种各样的选择。但是，现实中的管理人员所做的大多是管理团队和维持项目计划免遭终结等工作。如果他们组织各方协作完成一个项目，在最后时刻，预测结果突然变得起伏不定，而你却不能在最后关头放弃这一项目，不是吗？"

后来，汉森还说道："即使是专业学者，对收集预测跟踪记录也没有多大兴趣——他们对做出准确的预测不感兴趣。还有一个更为基本的问题，那就是我们的社会需要专家，但不怎么需要准确的预测。"

为了解决这一缺陷，汉森一直力挺那些预测市场—可以让人们在某一特定的经济或政策结果上投下赌注的一些系统，如以色列是否会与伊朗交火，或是气候变化会导致全球气温升高多少。对于这些问题，汉森的观点十分明确：预测市场能够保证我们在作预测时获得实实在在的经济利益，而不仅是在同行面前"赚面子"。

在本书后面的内容中，我们会再次对预测市场进行讨论。预测市场其实不是什么"万灵药"，尤其是在我们误认为它们无所不能的时候更是如此。但是正如汉森所言，预测市场至少可以通过整理人们的动机来改善预测质量。

预测市场最基本的应用可能，就是对 GDP 增长率和失业率等宏观经济变量进行预测。我们有各种直接或间接的方法预测通货膨胀、利率和物价波动，但是对于 GDP 增长率，却还没有大的预测市场。

或许也有人会关注这样的预测市场：近年来，普通股与宏观经济风险之间的联系越来越密切，因此，可以找到降低二者潜在风险的办法。

预测市场也可以为决策者提供实时信息，主要是不断更新的 GDP 增长率预测。人们对预测市场有了更多选择，如打赌 GDP 会增长 5%还是下降 2%，这对那些过度自信的预测者是一个不小的打击，而同时人们在预测经济时，也会得到更加可靠的关于不确定性的估值。

而另外一种解决方案——需求性选择要求我们必须成为更好的预测用户。在经济预测领域，这也许就意味着，我们应该把更多的目光投向像哈祖斯这样实实在在地探寻经济实质的人的身上，而不应关注那些将各种领先指标混在一起冒充预测专家的骗子；还可能意味着我们应该更加关注经济指标和经济预测中的噪声。与政治预测一样，GDP 增长率的初始预测值在公布时也应该提到其误差幅度。

从更广义的角度来看，在预测中表现出的自信并不能代表预测的准确度，相反，这两者经常成反比。在经济领域或其他领域中，当我们阻止预测者对我们周围世界的内在风险给予百分之百的重视时，就会埋下巨大的风险。

第七章

传染性疾病预测：禽流感为何会突然爆发，又突然消失？

每年 1 月，流感都会准时地在新泽西州迪克斯堡陆军基地蔓延，几乎成了一种惯例。绝大多数美国士兵每年都会回家过圣诞节，他们分散到全美各个角落休冬假。假期结束回到基地时，这些士兵个个膘肥体壮，精神饱满，但同时也把各自家乡盛行的流感病毒带回营队。若美国上下流感肆虐，那么士兵们很可能将病毒带回军营。军营的生活环境狭窄，毫无隐私和个人空间，如果某位患上流感的士兵回营，很容易就会传染给战友，几乎没有哪个地方比军营更适合病毒的传播了。

通常，这样的流感并不足以引起重视，每年一月和二月，美国都会有几千万人患上流感，但死亡病例却微乎其微。年轻健康的戴维·李维斯是一个罕见的例外，这位来自马萨诸塞州西艾希利的 19 岁的二等兵于 1976 年 1 月回到迪克斯堡后，不幸患病去世。当时队伍中李维斯病得最厉害，上级命令他留在营房休息，但他坚持同战友一起穿越新泽西中部白雪覆盖的沼泽区，完成 80 公里的行军。1976 年是美国建国两百周年，“水门事件”和越南战争之后，美国上下亟须良好的秩序和严明的纪律，在这样的大环境下，李维斯不想让发烧这点小事打乱自己的训练。

可不幸的是，李维斯再没能回到军营：行进到 21 公里处时，他突然倒下，随后被宣告死亡。尸检报告表明，李维斯肺部充血，死于急性肺炎，这是一种常见的流感并发症，通常不会导致李维斯这样健康的年轻人死亡。

而迪克斯堡的军医当时已经对那年的流感病毒高度紧张了。尽管数百名患病的士兵被检测出感染了甲型病毒维多利亚株（这是那年在全世界传播的一种常见的、危害不大的病毒），但还是有一些士兵和李维斯一样感染了一种显然严重得多的未知流感病毒。患者的血样已被送往亚特兰大的疾病控制中心作进一步检测。

两周后，疾病控制中心公布了这一种神秘病毒的类型，它并非新型病毒，而是令人更加不安的多种病毒混合体，这种由旧病毒衍生出的“恶魔”名为 H1N1 流感病毒。H1N1 病毒曾导致美国现代史上最严重的流行疾病：1918~1920 年的西班牙流感，这场流感导致全世界 1/3 的人感染，5 千万人丧生，其中包括 67.5 万名美国人。无论由于科学还是

迷信的原因，这些数据的公开都让美国流行病学界不寒而栗。1918 年流感最早的迹象也出现在一个军事基地，堪萨斯州的莱利堡基地，当时士兵正为参加第一次世界大战作准备。另外，1976 年时有一个观点——基于非常脆弱的科学证据——认为大约每 10 年就会爆发一次严重的流感。1938 年、1947 年、1957 年和 1968 年都发生过大流感，那么，1976 年应该也会爆发一次。

于是，骇人听闻的预测接踵而至。美国疾病控制中心确定这种新型病毒为 H1N1 后，并没有立刻引起公众的关注，因为流感季节已经结束。但很多科学家害怕接下来那个冬天会出现更可怕的情况。一位资深的医生告诉《纽约时报》，从未有哪种新型的流感病菌会败给它的对手，未能雄霸全球：但甲型病毒维多利亚株这样弱小的病毒根本就没有办法战胜更加厉害、聪明的对手。若此次 H1N1 病毒和 1918 年那次一样致命，后果将不堪设想。美国总统杰拉德·福特任期内的卫生部部长 F·戴维·马修预测会有 100 万名美国人死亡，1918 年的人口死亡总数与之相比真是“小巫见大巫”了。

福特总统在困境中挣扎着。像时尚产业一样，疫苗产业至少需要 6 个月的生产周期才能知道新疫苗对新一季的流感是否有效，这一过程每年的变动不大。突然要求这个产业生产对付 H1N1 的疫苗，尤其是产量还得足够供全美国使用，这个行业必须马上行动起来才行。同时，福特还得努力修正他在公众心中迟钝、不自信的形象——每周末热播的《周六夜现场》中切维·查斯的“恶搞”，使总统的这一形象越来越深入人心。所以，这次福特总统迈出了坚定的步子，他要求美国国会批准生产两亿支疫苗，并命令在全美推行大规模接种疫苗计划，这是继 20 世纪 50 年代乔纳斯·索尔克研发出小儿麻痹症疫苗后美国第一次大规模接种疫苗的行动。

新闻媒体将这次大规模接种疫苗行动描述成一场赌博。福特则认为这是在金钱和生命之间下注，而他选择了正确的一方。美国国会两院都以压倒性多数票通过了这几项耗资 1.8 亿美元的计划。

到了夏天，人们对美国政府的计划产生了强烈质疑。尽管美国的夏季是流感低发期，但此时的南半球正处在冬季，通常这个季节是那里的流感高发期。但从新西兰的奥克兰到中美洲的阿根廷，都毫无 H1N1 病毒的踪迹，相反，温和常见的甲型病毒维多利亚株卷土重来，再次成为主要病菌。的确，全球已确诊的 H1N1 就只有发生在迪克斯堡的约 200 个病例，死亡的只有二等兵李维斯一例。于是，指责和批评从四面八方纷涌而来：从美国疾病控制中心的副主任到世界卫生组织，从著名的英国医学期刊《柳叶刀》到《纽约时报》的社论专版（该专版那时就认为

H1N1 的威胁是一个“错误警报”），到处都在指责这项计划。其他西方国家并没有像美国这样采取激烈的手段应对这种新型病毒。

然而，福特政府并不承认他们高估了这次流感的威胁，相反，他们还“变本加厉”，准备了一系列让人心惊肉跳的公益广告，当年秋天在美国国内的多家电视台定期滚动播放。其中一条广告讽刺了那些拒绝注射疫苗的人的天真无知，画面中一个谢顶的普通男子说:“在你见过的 55 岁的人当中，我是最健康的，每周末我都打高尔夫球。”然后，镜头便切换成几天后他在病榻上弥留之际的画面，这就是拒绝接种疫苗的结果。还有一则广告，一位女讲解员在缓缓地向人们展示病毒的传播途径，那种小心翼翼的口吻让人觉得这种新型病毒仿佛是某种性传染病——“贝蒂的妈妈把病毒传播给了出租车司机……司机又传染给一位迷人的空姐……她又将病毒传染给朋友多蒂，多蒂有心脏病，病毒夺去了她的生命。”

这些滑稽可笑的公益广告旨在传递一条非常严肃的讯息：要心存畏惧！美国民众对此心领神会，然而，他们对这种疫苗的恐惧丝毫不亚于对这种疾病本身的恐惧。纵观美国历史，凡美国政府实施全面注射疫苗时总会引起过度恐慌。但这次公众的疑虑似乎是有理有据的。1976 年 8 月，由于多家医药公司施压，美国国会和白宫同意，万一产品出现制造缺陷，可免除医药公司的法律责任。政府部门的这一决定被广泛解读为缺乏自信的做法，好像当时正忙着赶制疫苗，根本就没有足够的时间对其进行检测。那个夏天的民调显示，只有约 50%的美国人打算接种疫苗，远低于政府设定的接种率为 80%的目标。

到了 1976 年 10 月，疫苗接种计划正式开始，民众的不安情绪也随之达到最高点。10 月 11 日，匹兹堡报出有 3 位老人接种流感疫苗后不久便去世了，俄克拉荷马也报出两名老人注射疫苗后死亡，劳德岱堡也有一例类似状况发生。但是，没有证据显示这些死亡和疫苗接种有直接关系，毕竟每天都有老人离世。但是，由于人们对政府疫苗计划心存顾虑，加之媒体对这些统计数据居心叵测的理解，每报出一个与流感疫苗有关的死亡案例，都会引发民众的惊恐，甚至连全美最受信任的沃尔特·克朗凯特都一改其谨言慎行的常态，劝诫媒体不要再用那些耸人听闻的方式作新闻了，然而这些都无法让民众冷静下来。匹兹堡和其他很多城市已经关闭了疫苗注射诊所。

到秋末时，又出现了另外一个新的问题，这个问题比上次（老人注射完疫苗之后死去）的问题严重得多。大约有 500 名病人在接种完疫苗后开始出现吉兰·巴雷综合征的病症，吉兰·巴雷综合征是一种罕见的神经系统疾病，会引发瘫痪的自身免疫性疾病。这一次的统计数据更具说

服力：在普通人中，吉兰·巴雷综合征的发病率只有百万分之一，而接种疫苗的人患病概率是普通人的10倍，也就是说5 000万人中就有500人患病。尽管科学家还未查明接种疫苗会引起吉兰·巴雷综合征的原因，但是紧急的疫苗生产计划引发的制造性缺陷似乎成了合理的指控对象，医学界一致认为应该永远停止接种疫苗计划，但直到12月16日美国政府才下令停止。

最终，迪克斯堡爆发的H1N1病毒被完全隔离，之后美国再无确诊病例。同时，普通流感病毒甲型病毒维多利亚株导致的死亡人数也低于1976~1977年冬季各类事故死亡人数的平均值。一切针对H1N1采取的措施似乎都是无事生非。

很快，人们就给这次事件起了一个名称——H1N1惨败记。对于当时的福特总统而言，这次事件就是一场彻头彻尾的灾难，在1976年11月的美国总统大选中，福特败给了民主党成员吉米·卡特。那些药品生产商此前已经获得免责权，将超过26亿美元的巨额赔偿责任完全推给了美国政府。而且，几乎每家地方报纸都会刊登这类故事：某位贫穷的女服务生或学校教师为了履行义务，去接种了疫苗，却不幸患上了吉兰·巴雷综合征。随后两三年，愿意接种流感疫苗的美国民众的数量缩减，人数降到了100万左右，这样一来，如果1978年或1979年发生了一次大规模的流感，美国就非常有可能陷入极端危险的状况中。

当时的美国总统福特对H1N1新型流感病毒的处理，从很多方面来说都是不负责任的。福特以1918年的流感类型来推测此次流感病毒爆发的可能性，完全无视医学专家的建议。当时的专家认为，如此严重的情况发生的概率不可能超过35%，且极有可能只有2%。

究竟是什么原因导致H1N1新型病毒突然爆发，随后又突然消失？这个问题至今仍未明确。33年后，H1N1病毒再次爆发，这一次关于该病毒的预测并没有比之前准确。2009年H1N1流感爆发时，科学家们一开始并没有识别出这种病毒，而在检测结果出来之后，他们又过高地估计了这一病毒的危害。

人们又一次高估了H1N1的致命率

流感病毒借助鸟类这一寄主永生不灭，尤其是信天翁、海鸥、野鸭、天鹅和大雁这类海洋鸟类，它们带着病毒的基因在不同大陆之间迁徙，这些鸟很少染上疾病，却把病毒传播给其他物种，尤其是猪和鸡这类人类在日常生活中随处可见的家禽。鸡患上流感后，往往能战胜病毒存活下来，却会把病毒传给饲养者。猪的情况也是如此，因为它们不仅容易染上人类或是鸟类的病毒，其自身还会产生病毒，猪为不同种类的

病毒混杂变种提供了孕育的温床。

某一区域若满足以下 3 个条件，就会成为 H1N1 病毒的完美孵化器：

1.人、猪毗近，也就是说，猪肉是日常饮食中的主食。

2.靠近海洋，猪和越海迁徙的鸟类可能发生接触。

3.很可能位于发展中国家，由于国家贫穷、个人卫生和公共卫生水平较低，动物病毒更易传播给人类。

上述 3 个条件正好描绘出东南亚许多国家的现状，比如中国、印度尼西亚、泰国和越南（仅中国现存猪的数量就占世界数量的 1/2）等国。这些国家通常都是流感的源头，从每年常见的流感类型到特殊的变体，而后者可能引起全球性的流行病。所以，这些国家成了医学界关注的对象，尤其是近些年，由于担忧新型病毒会爆发，医学界更加关注这些国家了。H5N1 病毒，俗称禽流感，已经在东南亚地区酝酿了许多年，一旦发生突变，极其容易致命。

然而，这些情况并非亚洲独有，世界上其他地区也有类似情况，比如墨西哥的维拉克鲁斯。维拉克鲁斯毗邻墨西哥湾，同为发展中国家的墨西哥，传统饮食中猪肉的比例很大。此前几乎没有科学家在维拉克鲁斯探寻流感的迹象，但是，2009 年 H1N1 病毒就是在这里爆发的。

2009 年 4 月底，科学家收到很多关于维拉克鲁斯和墨西哥其他地区 H1N1 疫情的数据统计。墨西哥共报出 1 900 例 H1N1 的病例，死亡数量达 150 例，死亡率达到 8%。这个比率已经超过了西班牙流感的致死率。而且，死亡病患多数为年轻、健康的成年人，这是重大疾病爆发的一大特征。另外，该病毒还具有很强的再生能力，除了墨西哥和美国，在加拿大、西班牙、秘鲁、英国、以色列、新西兰、德国、荷兰、瑞典和冰岛等国也都检测出该病例。

H1N1 病毒顷刻之间（而不是 H5N1）成了令科学家一直以来心存恐惧的超级病毒。整个墨西哥城基本上都被隔离了，许多欧洲国家都提醒其公民不要到墨西哥或美国旅行。中国香港和新加坡因对流感蔓延产生的恐慌情绪，令股市暴跌。

但人们最初的担忧很快就平息下来。尽管 H1N1 病毒在美国传播的速度确实很快——2009 年 4 月 26 日确诊了 20 例，15 天以后患病人数就迅速升至 2 618 例。出人意料的是，大多数病例症状都属于轻度流感，确诊的死亡病例仅为 3 例，与一般的季节性流感危害类似。一周

后，美国疾病控制中心就建议学校重新开放。

然而，这种疾病仍在全球继续蔓延，2009 年 6 月，世界卫生组织发布了 6 级流行病预警，这也是最高级别的预警。科学家们担心这次流感会像 1918 年西班牙流感一样，起初患病时症状轻微，但第 2 波和第 3 波发病潮到来时就会变得更加致命（见图 7–1）。2009 年 8 月，美国政府称可能会有 1/2 的美国人感染 H1N1 病毒，死亡人数约为 9 万人，这一貌似可信的消息使美国民众的情绪再次悲观起来。

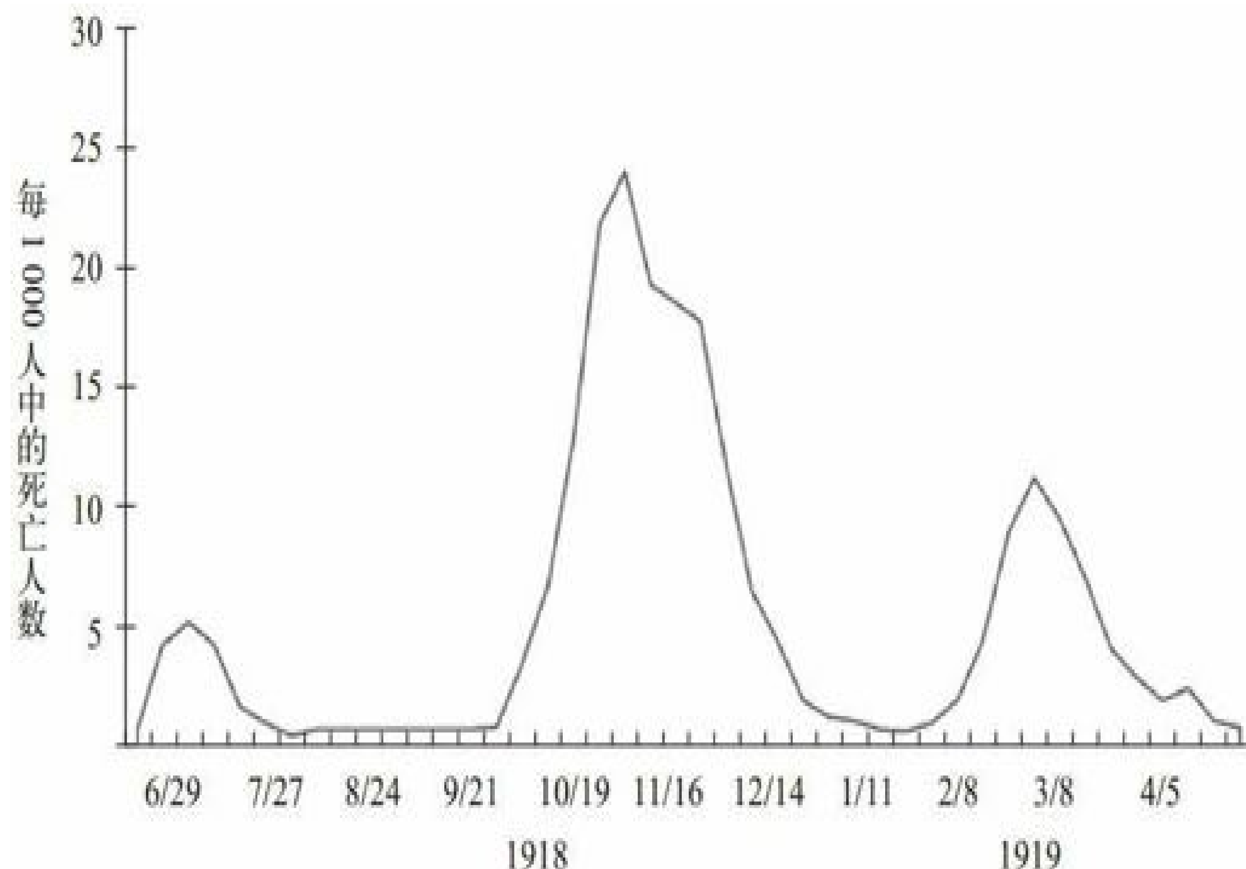


图 **7–1 1918~1919** 年 **H1N1** 流感疫情造成的死亡率

然而，这些预测也都被证实是毫无根据的。美国政府最终报告 2009 年感染 H1N1 病毒的人数为 5.5 千万人，占全美总人口数量的 1/6，而并非之前认为的 1/2，死亡人数为 11 000 人。其实，H1N1 病毒致死率仅为 0.02%，并不属于异常严重的病毒，而是属于格外温和的病毒。实际上，2009 年流感病毒的死亡人数比因常见的季节性流感而死亡的人数还要少。这次预测并不像 1976 那次那样尴尬，但由始至终预测结果都不够准确。

人们无法保证下一次流感预测会更准确。事实上，流感和其他传染病的某些特征本身就决定了对其预测将会十分困难，并充满挑战。

外推法的危害——艾滋病感染人数被低估了一半

外推法是一种非常初级的预测方法。这一方法仅仅包含一个假说，即未来是现在趋势持续发展的结果。众所周知的预测失败的案例中，有些就是由于太过随意地应用了这一假设。

例如，19 世纪末 20 世纪初，许多城市的规划者被马粪所扰，马车的数量日益增加，使得路上的马粪越积越多。1894 年一位来自《伦敦时报》的作家致力于研究街道上的马粪问题，他预测到 20 世纪 40 年代，伦敦每条街道会被厚达 2.7 米的马粪覆盖。幸运的是，大约 10 年后，亨利·福特生产了福特 T 型汽车，避免了这场马粪危机。

外推法也是人口预测失败的罪魁祸首。1682 年，英国经济学家威廉·佩蒂爵士作了一项针对全球人口数量增长的预测，这项预测也许是第一次正式对人口数量进行的预测。由于当时无法获得丰富的人口统计数据，威廉便采取了大量颇具创新性的方法对人口增长进行推断，相当准确地预测到全球人口在 17 世纪增长速度将放缓。然而，威廉却错误地假设人口增长会一直保持这样的趋势，到 2012 年全球人口数量可能将过 7 亿。一个世纪之后，工业革命爆发，人口数量增长率直线上升。世界人口数量在 2011 年年末已超过 70 亿，约为佩蒂预测结果的 10 倍。

斯坦福大学生物学家保罗·R·埃利希和妻子安妮·埃利希于 1968 年出版了一本颇具争议的书，名为《人口爆炸》，这本书犯了相反的错误，错误地预测 20 世纪 70 年代会有一场饥荒夺去几亿人的生命。这项预测之所以失败，原因有很多，包括埃利希夫妇一直倾向的末日说，他们对引发末日的蛛丝马迹颇为关注。但主要问题是，埃利希夫妇认为 20 世纪 60 年代性解放时期创纪录的高生育率会无限期地延续下去，需要抚养的人口会越来越多。“在写《人口爆炸》一书时，我认为人类对性爱和孩子的兴趣浓厚，家庭人口很难减少。如果公正有礼地对待女性，给她们提供就业机会，生育率自然会下降。”在一次简短的采访中，保罗这样对我说。那些没有作出这种过分简单化假设的学者当时就意识到这一点了，一般联合国针对 20 世纪六七十年代发布的人口规划方案的预测都很准确，在这一规划方案中提到未来三四十年世界人口的状况。

在研究对象的数量总在急速增加的领域——包括人口数量增长及疾病领域——中使用外推法常会引发严重的问题。20 世纪 80 年代初期，美国的艾滋病患者数量呈指数型增长：整个 1980 年共出现 99 例艾滋病患者，1981 年又出现 434 例艾滋病患者，1984 年出现的艾滋病患者竟高达 11 148 例。你可以将这些数据画成图并据此推断未来的发展模式（当时一些学者就是这么做的），便会得出这样的预测结果——到

1995 年美国确诊的艾滋病患者人数会升至约 270 000 人。这一预测本来就不是很准确，可不幸的是，预测结果低估了艾滋病毒的危害性。1995 年确诊的艾滋病病例约为 560 000 人，比预测总数高出一倍。

然而，从统计学的角度来看，更大的问题也许是预测对象的数量呈指数型增长时，用外推法均无法做出准确预测。若恰当使用外推法，就会考虑到这一方法的误差幅度，那么得出的结果就应该是 1995 年美国感染艾滋病的病例数可能低到 35 000 例，也可能高达 180 万例。以预测观点来看，这么宽泛的幅度并不能说明任何问题。

为什么 2009 年的流感预测会失败？

在流感爆发的初期，流行病专家使用的统计方法并不会如前文所提到的那样简单，但是这些专家也可能会使用外推法，根据不可靠的数据进行预测。

一种名为基本传染数的变量是预测疾病蔓延最有用的变量之一，通常标记为 R0，这一变量可以测量一个感染者将病毒传播给未感染者的可能数量。比如 R0=4，指在没有采取疫苗注射和其他预防措施时，一个感染者在康复（或病亡）前会将病毒传播给 4 个人。

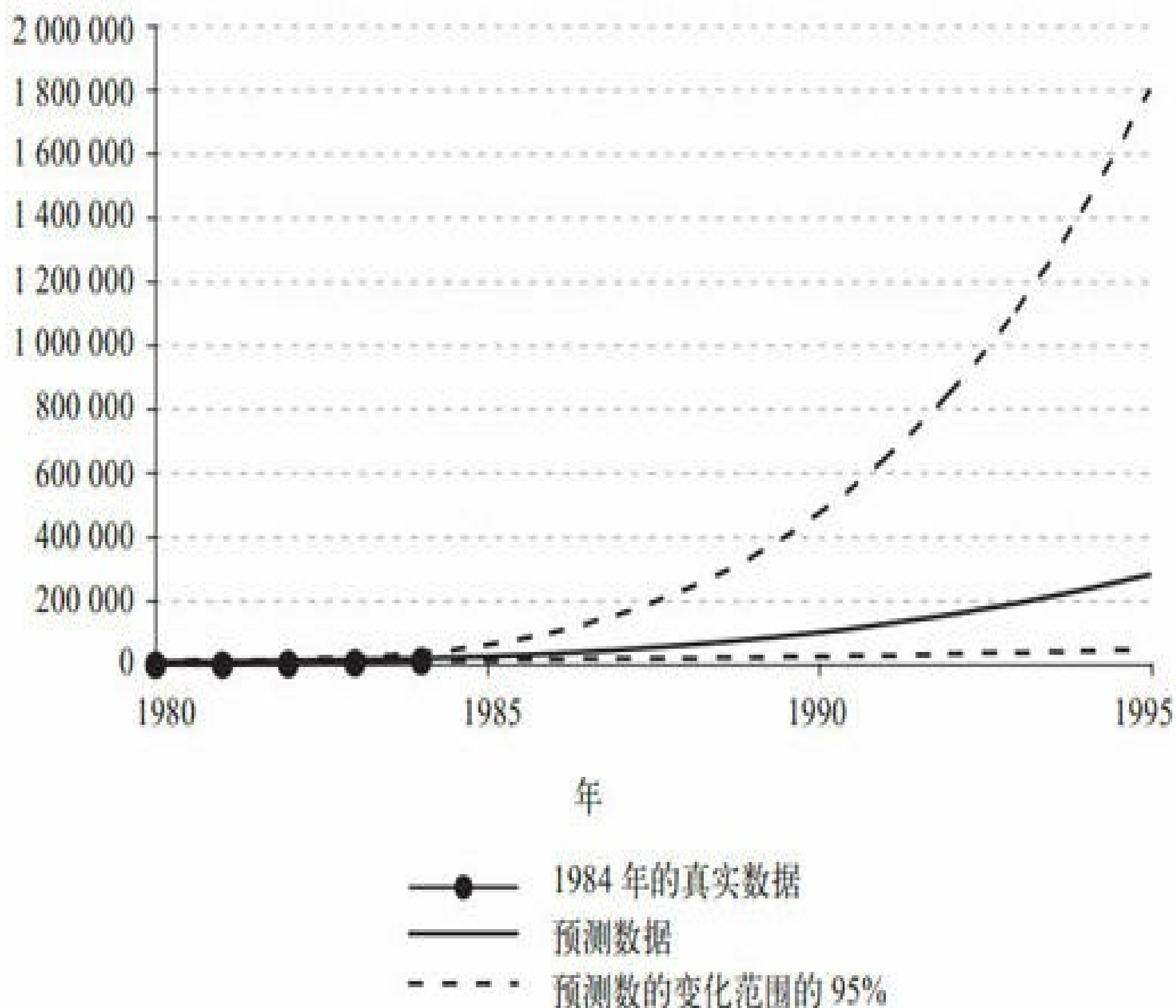


图 **7–2** 美国累计确诊的艾滋病人数：**1984~1995** 年确切统计的数据和根据外推法预测数据

从理论上讲，在没有注射疫苗或采取隔离手段干预的情况下，任何一种 Ro 大于 1 的疾病最终都会使全人类染病。但有时 Ro 的数值会远大于 1：西班牙流感的 Ro 值为 3，水痘的 Ro 值为 6，而麻疹的 Ro 值则高达 15。作为人类文明史上致死率最高的疾病之一，疟疾的 Ro 值可能会达到 3 位数，目前，世界某些地区仍有 10%的人死于疟疾。

问题是，可靠的 Ro 估值总是要等到该疾病在某个群体中迅速传播开来之后才能明确地计算出来，得有足够的时间仔细检查统计数据才行。所以，流行病学家不得已只好依据少量的早期数据来作外推预测。另外一个关键的疾病统计量是致死率，这一数据在疾病爆发初期也很难准确预测。这就好比“第二十二条军规”，会使人陷入两难境地，但是没有这一数量，就很难准确地预测某种疾病，要对这些未知量作出可靠预估，只有等到疫情结束。

表 **7–1** 不同疾病的 **Ro** 值的中位数估值

疟疾	150
麻疹	15
天花	6
艾滋病	3.5
非典	3.5
1918 年 H1N1 流感	3
1995 年埃博拉病毒	1.8
2009 年甲型 H1N1 病毒	1.5
季节性流感	1.3

而且，一种传染病爆发之初的数据经常被误报。例如，之前列举的美国艾滋病诊断病例的数据是在事件发生多年后才统计出来的。即使是时时更新的数据，对预测的帮助也不大。然而，如果依据科学家们当时实际使用的数据，预测结果会更糟糕。这是因为艾滋病出现之初，很多病人（甚至很多医生有时）对这种疾病所知甚少，使预测更加困难。许多带有艾滋病症状的奇怪并发症尚未查明病因或被误诊——艾滋病引发的机会性感染一直被误认为是这种疾病的主要死因。几年后，医生重新审阅旧的病例记录，才对艾滋病出现之初的发病率作出比较准确的估值。

不准确的数据也是导致 2009 年 H1N1 病毒预测失败的原因之一。那一年，墨西哥 H1N1 流感病毒的致死率出奇的高，而美国 H1N1 流感病毒的致死率却极低。尽管在某种程度上，这与两国医疗服务水平的差距有关，但主要差异还是统计假象。

致死率是相对简单的比率关系：由病死人数除以患病人数得出。但是，方程式中的两个数据都有较大的不确定性。一方面，墨西哥倾向于把其他形式的流感甚至其他疾病导致的死亡全都归因于 H1N1。实验测试显示，被认定为 H1N1 致死的患者中有 1/4 的人表现出完全不同的病症。另一方面，H1N1 流感病例数肯定存在少报、漏报的情况，也许报告值和实际值之间相差几个数量级。像墨西哥这样的发展中国家，既没

有美国那种成熟的新闻报道体系，也没有“一有不适便去就医”的生活习惯。H1N1 病毒进入美国之后的传播如此之快，可想而知，在墨西哥很有可能有几千甚至几万个轻微甲流病例没有上报政府。

事实上，H1N1 病毒已经在墨西哥南部和中部地区传播了很久，数月后才引起医学界的关注（尤其是在医学专家开始忙于关注亚洲的禽流感之后）。2009 年 3 月初，有报道称维拉克鲁斯市一个名叫洛里亚的小镇爆发呼吸系统疾病，当时镇上大部分人都已染病，但是墨西哥政府最初认为该病是由一种更为常见的流感病毒 H3N2 引起的。

与之形成鲜明对比的是，H1N1 病毒一进入美国就成为媒体炒作的噱头，几乎没有什么病例可以逃过媒体记者的眼睛。有了如此高质量的报道，美国因 H1N1 病毒致死的数据想必是相当可靠的。即使后来将一些最糟糕案例的情况从报道中拿掉，但已经太迟了，它们已经引发了公众的极大恐慌。

自我实现预测与自我否定预测

在许多涉及人类活动预测的案例中，预测行为本身就会改变人的行为方式。有时，这些行为的改变也会对预测结果产生影响，要么使结果无效，要么令结果更加准确，在经济学领域就是如此。流感和其他传染疾病的预测也受到这一问题的两方面影响。

若预测会自动实现，这种情况便可称为自我实现预测，或自我实现预言。在一场多候选人的竞选中，如美国总统初选，随着政治民调的公布，这类预测情况就会出现。在这类竞选中，投票者会战略性地选择那些有可能获胜的候选人，这样就不会浪费自己的选票了，此时，一份公之于众的民意调查就会成为候选人成功与否的最佳风向标。举个例子，2012 年艾奥瓦州共和党核心竞选的后期，美国有线电视新闻网（CNN）发布的一项民调显示，瑞克·桑德鲁的支持率奋起直追，拥有 16%的投票率，而之前他的支持率只有 10%。这份民调与众不同——在 CNN 公布这份民意调查之前，其他的调查均未显示桑德鲁有奋起直追之势。这份民调为桑德鲁营造了良好的媒体宣传氛围，一些选民开始放弃那些执政理念与其相似的候选人，如迈克·贝奇曼和瑞克·佩里，转而把选票投给桑德鲁。不久之后，桑德鲁赢得了艾奥瓦州的选举，而贝奇曼和佩里早就结束了竞选。

设计和娱乐等领域有许多更加微妙的例子。这些领域间的竞争主要是猜测消费者的喜好——但是，商家也可以通过巧妙的营销计划影响消费者的喜好。在时尚界，有种类似家庭手工业的机构会对下一季的流行色进行预测——通常要提前约一年作出预测，因为制衣生产线的调整需

要时间。如果一些有影响力的设计师将棕色定为来年的流行色，并开始大量生产棕色的衣服，还让模特和名人穿着棕色服装，门店的橱窗和宣称小册子上也都主推棕色，那么，公众可能也会逐渐追随这一潮流。而公众的反应往往是购买和消费棕色的商品，不仅仅是口头表达一下内心偏好。于是，“预测”到流行色的设计师在人们眼里便如同圣人一般，但即便当初设计师没有选择棕色，而是选了白色或黑色或淡紫色，情况也都是一样的。

同样的，疾病和其他身体状况也有这种自我实现的特性。当媒体广泛地讨论某些疾病时，人们就更有可能辨识出这些疾病的症状，医生也更有可能对这些疾病做出诊断（或误诊）。近年来，最有代表性的案例就算孤独症了。若将被诊断患有孤独症的儿童数量和“孤独症”一词在美国报纸中出现的频率相比较，你会发现两者旗鼓相当，几乎分毫不差（见图 7–3），近年来这两项数字明显都有所增加。人们并没有正确地看待孤独症这种疾病，这一疾病如今大有与流感平分秋色之势。

哈佛大学公共卫生学院的艾利克斯·欧祖诺夫博士告诉我：“这是一个很令人着迷的现象。在没有因果关系的疾病中，是新闻事件促使报告量不断增加。”欧祖诺夫接受过纯粹的数学训练，在许多数据驱动的领域也是如鱼得水，但现在他正潜心研究如何在流感和其他传染性疾病中运用严格的统计分析。“我们一而再，再而三地发现，人们越担心某种特定情况，这种情况就越容易成为当下讨论的话题，相关报道就越接近真相。”

欧祖诺夫认为这一现象也许可以解释 2009 年 H1N1 病毒席卷美国时的迅速。这一疾病的传播速度确实很快，而有些人将原本可以忽略不计的病症也都一一报告给他们的医生，于是统计数据激增。

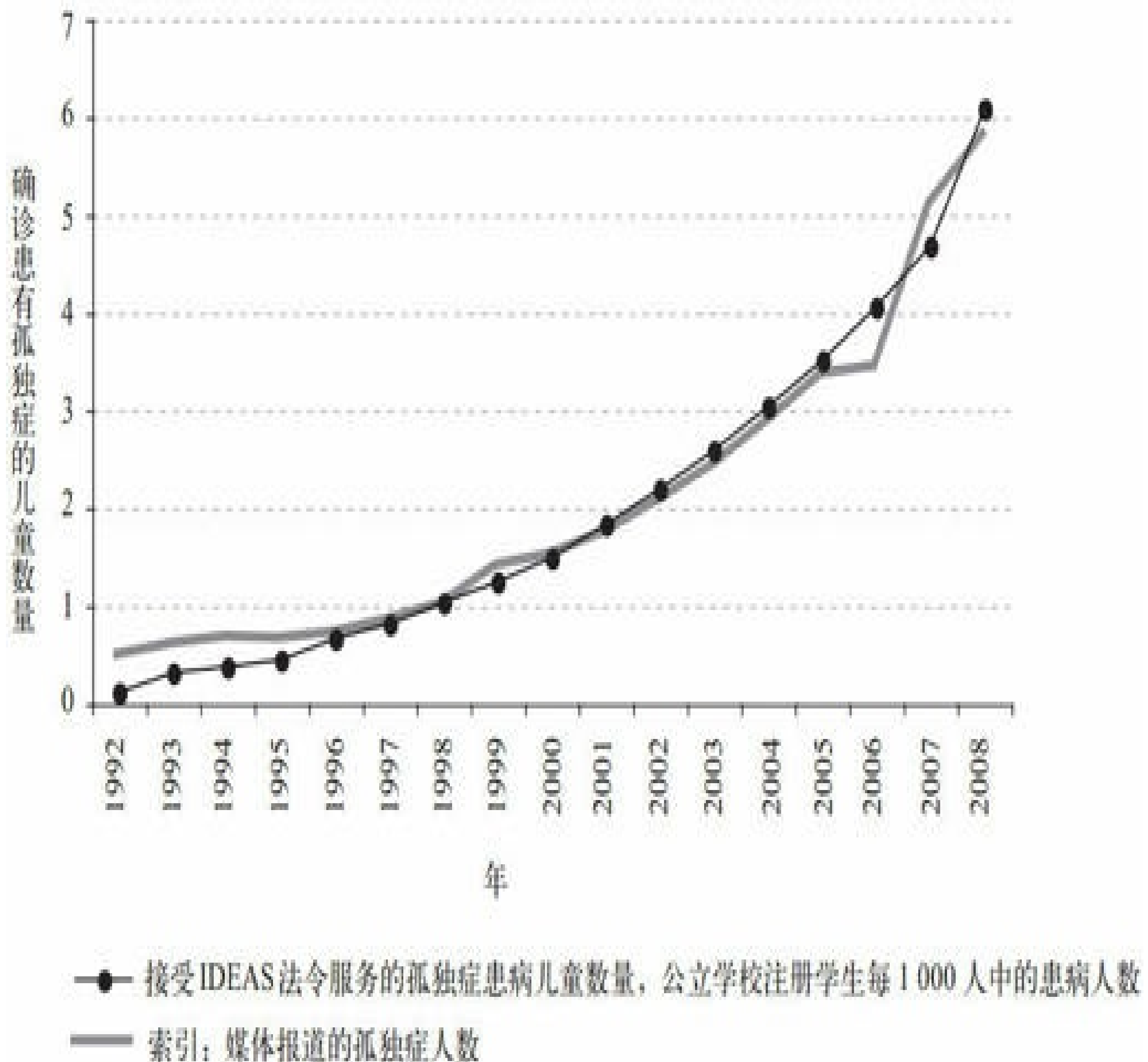


图 **7–3** 孤独症：**1992~2008** 年媒体报道病例与确诊病例

如果医生想要预测疾病在人群中的发病率，公开报道的病例数量会误导这一评估。这与犯罪报道的情况相类似：如果警察报告某一居民区的偷盗案件数量增加，是因为警察变得更警惕了，抓到了之前遗漏的罪犯，还是因为这类案件的报道更随意了呢？或者因为这一居民区变得更危险？任何想在流感发生初期做出预测的人都会碰到这类让人感到困惑的问题。

与自我实现预测相反的是自我否定预测，自我否定预测是指预测会自我破坏。越来越普及的 GPS 就是一个有趣的例子。曼哈顿有两条南北走向的主干道，一条是西部高速公路，靠近哈德逊河；另一条是罗斯福路，位于曼哈顿东部。根据目的地的位置，司机也许没有必须走哪条

路的强烈愿望。然而，GPS 导航系统会依据车流量预测哪条路通行的时间更短，然后用语音提示你应该选的道路。可是，当很多车主都用同一款导航仪时，问题就出现了，大家都会选同一条路，于是道路突然会被车流塞满，畅通的路反而变得拥堵。理论和现实的双重证据表明，在纽约、波士顿和伦敦都出现了类似的问题，这些导航系统的作用有时只会适得其反。

因为流感预测的目标从某种程度上来说是为了提高公众对这种疾病的意识，进而改变其行为习惯，因此这种自我破坏的特性也影响了流感预测的准确性。最有效的流感预测很可能就是无法修成正果的预测，因为这样的预测会促使人们做出更有利于健康的选择。

预测模型越简单越好，还是越复杂越好？

芬兰科学家汉娜·库克将构建统计或预测模型比作绘制地图。绘图需要足够多的细节才能真实地展现基本景观，于是，制图者不愿漏掉大城市、主要河流、山脉，或是重要的高速公路。然而，太多细节会让旅行者晕头转向，反而迷了路。正如本书前文中提到的那样，这些问题不是纯粹的审美问题。过于复杂的模型可能会将噪声拟合进来，无法成功地复制内在结构，使预测失败。

但是，多少细节才算太多（或是太少）呢？地图绘制技术需要穷尽一生才能掌握，这项工作是艺术和科学的有机结合。将模型的建立描述成艺术形式也许有些离谱，但这项工作确实需要作很多判断。

然而，从理想化的角度来看，像库克提到的那些问题也可以凭经验回答。这样的模型有效吗？如果没有，就需要寻求其他的解决路径。在流行病学界，医生所用的传统模型相当简单，效果也没有那么理想。

传染疾病最基本的数学处理模型就是 SIR 模型（见图 7–4）。该模型形成于 1927 年，假定某人在某时可能会处于以下 3 种状态中的一种：S 代表易受疾病传染，I 代表被感染，R 代表康复。像流感这类不算严重的疾病，从一种状态转变到另一种状态完全是单向的：从 S 到 I，再到 R。在这一模型中，疫苗基本上充当了捷径的角色，可以让某人直接从 S 到 R，免去生病的过程。这个模型背后的数学过程是相对直接的，（将题目）概括成不同的方程式，在电脑上经过几秒钟的运算就可以得出结果。

问题是，这个模型要有很多假设才可以有效运转，而其中有些假设在实践中根本无法实现。尤其是这个模型假设在某个特定人群中，每个人的行为方式都是一样的。他们同样易受传染，都可能接种疫苗，彼此随意接触；他们中并没有种族、性别、年龄、宗教信仰、性取向以及教

派的区别；每个人的行为方式基本上没有什么区别。

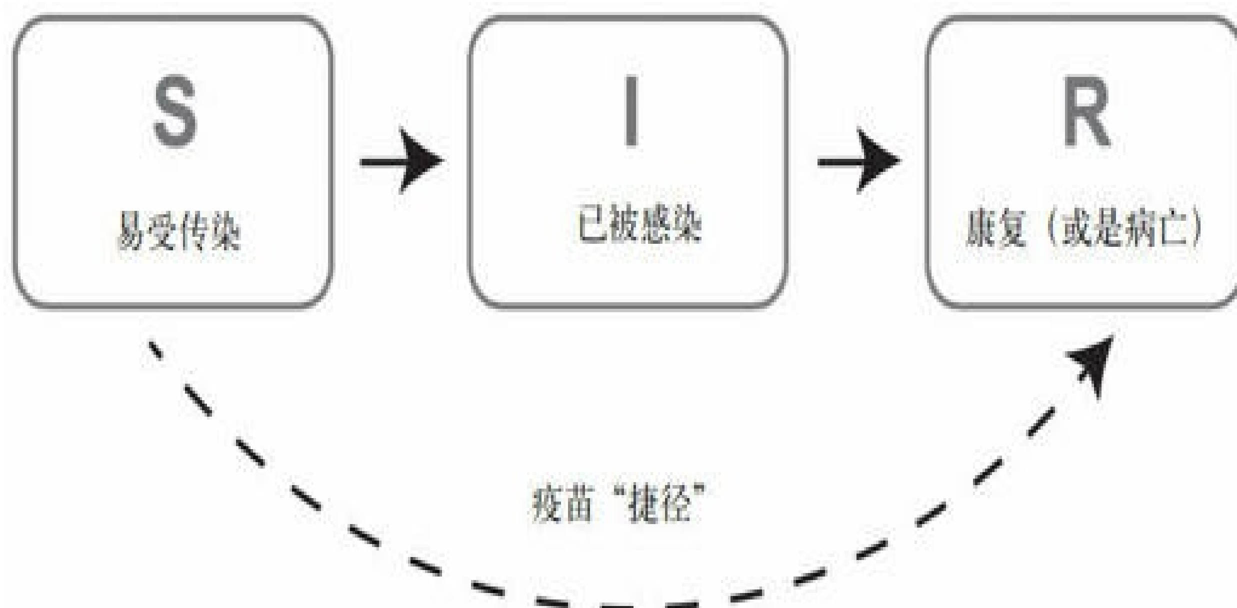


图 **7-4 SIR** 模型的示意图

在那些通过性行为传播的疾病中，最容易看出这些假设的缺陷。

20 世纪 90 年代后期到 21 世纪初，旧金山的男同性恋群体中无保护措施的性行为越来越普遍，而早在 20 年前这群人就曾受到人类免疫缺陷病毒/艾滋病病毒的侵蚀。一些研究者将这一情况归责于吸毒人数的上升，尤其是吸食冰毒常会引发更危险的性行为。其他研究人员认为是由于抗病毒治疗的效果越来越好——“鸡尾酒治疗法”可以让人类免疫缺陷病毒携带者的寿命延长几年甚至十几年：对男同性恋者而言，HIV 不再等同于死刑判决。然而，其他理论则比较关注时代模式——20 世纪 80 年代艾滋病盛行的旧金山，在新一代男同性恋群体眼中似乎已成为一段遥远的历史了。

但专家们都同意这样一种说法，那就是随着无防护措施的性行为逐渐增多，HIV 病毒感染率也会随之升高。

但是，结果证明 HIV 病毒感染率并未升高，其他性传播疾病反倒增加了：在男同性恋者中（MSM），一种新型的梅毒——20 世纪 90 年代，梅毒在旧金山已基本根除了——又开始快速传播，病患人数从 1998 年的 9 例增至 2004 年的 502 例。淋病发病率也有所增加。然而，矛盾的是艾滋病病例却并没有增加。2004 年,梅毒发病率达到多年来的最高值，而 HIV 确诊数却降至继艾滋病传播以来的最低值，对于研究者来说，这种现象并不合理；梅毒和 HIV 病毒在统计学上通常紧密相关，存在着因果关系，因为一旦染上其中一种病毒，就会更容易染上另外一种。

现在看来，这个悖论已经有了解决方法，那就是男同性恋者“血清分类”的效果越来越好，也就是说，男同性恋者在选择性伴侣时，会选择和自己 HIV 情况相同的人。他们是如何做到这一点的呢？对此众说纷纭，但在旧金山、悉尼、伦敦和其他同性恋人口居住较多的城市，对大量男同性恋者所作的详细行为研究中的确记载了这种做法。或许是因为，公共卫生宣传发挥了一些积极作用，有些宣传对“避孕套疲劳”避而不谈，而是强调“协商安全”的理念。从某种程度上说，也许互联网已经取代同性恋酒吧，成为同性恋者挑选性伴侣的首选之地。互联网公开信息要遵守不同规定：许多人都会在基本资料中填写 HIV 状况，另外，在家中的私密空间问一些平时难以启齿的问题要比在酒吧喧闹的舞池里问得更容易一些（也更容易得到诚实的回答）。

无论原因是什么，显然，这类具体的、局部的行为混淆了更为简单的疾病传播模式——幸运的是，这种情况意味着这些模型高估了 HIV 病毒。SIR 这类阶段模型认为每位个体都容易受疾病传染。可是，当某些疾病是通过十分亲密的接触传播的，或当不同种类群体染病的风险是不一样的，SIR 模型就行不通了。你不可能随便走进一家商店，回家时就携带了 HIV 病毒。

然而，即使是某些更简单的疾病，这种阶段模型也会因为其假设过于宽松而失败。以麻疹为例，绝大多数刚入行的流行病学家，在读博士阶段的第一个研究对象就是麻疹，因为这种病最容易研究。欧祖诺在哈佛大学的一位同事马克·利普思奇称：“麻疹是传染病的模型系统。只需进行一个血液测试，就能很明确地判断出这种疾病，因为麻疹病毒只有一种类型，所有带有这种类型的病毒的患者都是麻疹患者。一旦患过麻疹病，就不会再次发病。”如果说有哪种疾病能采用 SIR 模式来处理，那恐怕就只有麻疹了。

20 世纪 80 年代到 90 年代初，芝加哥爆发了严重的麻疹疫情，流行病学家在预测时感觉困难重重。传统模型显示，已经有足够多的芝加哥居民接种了麻疹疫苗，这些人本可以营造一种“群体免疫效应”——从生物学角度来看，这种效应就相当于“防火墙”，接种了疫苗之后疾病就没有机会传播，渐渐就会消失。然而，20 世纪 80 年代，几年内有近千名芝加哥人患上了麻疹，其中大多数的患者是孩子；情况十分严重，该市下令让护士挨家挨户为大家注射疫苗。

来自芝加哥大学医学院的儿科及传染病专家罗伯特·多姆博士对这次麻疹爆发作了深入研究。多姆医术高明、声音厚重、络腮胡子、有一种超然的幽默感。我采访多姆和他在芝加哥大学的两位同事时，多姆刚从海地回来，他参与了 2010 年的海地地震救灾工作。

我曾在芝加哥住过 13 年，那是一座由一片片社区组成的城市。这些社区通常都是彼此分离，不同种族、不同社会经济地位的人很少混居在一起。多姆发现这些社区对疫苗接种的倾向也不同：内城居民大多比较穷困，南部的黑人很少让孩子接种 MMR（麻疹、流行性腮腺炎、风疹）疫苗。那些未接种疫苗的孩子一起上学，一起玩耍，打喷嚏都是面对面。这些人违背了 SIR 模型称为随机组合的假说，该假说认为人群中任意两个人彼此接触的可能性是一样的。于是，麻疹在黑人孩子中传播。

这个非随机组合的现象也是 1976 年 H1N1 病毒惨败记的诱因，当时的科学家试图根据迪克斯堡出现的病例外推出美国受 H1N1 病毒危害，而迪克斯堡就是一个非随机组合。从局部来看，那次 H1N1——现称为新的流感毒株 A/新泽西/76——十分危险，因为在军事基地中，该病毒传播的速度非常快，两三周内就有确诊病例 230 例。科学家由此推断该病毒的基本再生率肯定相当高——或许和 1918 年大流感的 R0 值一样，约为 3。

然而，军队的环境向来易于传播疾病。士兵们彼此接触密切，在军营里常共享食物和床上用品，几乎没有私人空间。另外，士兵们还常常进行高负荷的体能训练——这会使他们的免疫系统暂时衰竭——而社会规范认为军人即使生病了也得坚守岗位。于是，传染病就有无数的机会在士兵中传播，传播速度通常会非常快。

随后对迪克斯堡爆发的流感研究显示，疾病的快速传播是由环境因素引起的，与疾病本身的危害程度没有直接关系。迪克斯堡完全不同于任何美国社区或工作场所。事实上，根本不必担心 A/新泽西/76 流感，其 R0 值仅为 1.2，它与一般季节性流感的情况相类似。在军营外，或者在除了类似军营环境的大学宿舍、监狱之外，流感并没有那么严重。这种流感基本上在迪克斯堡已经消失了，不会传给新的个体。

A/新泽西/76 流感的传播以彻底失败告终，就像旧金山 HIV 或梅毒的悖论，或是 20 世纪 80 年代的芝加哥麻疹疫情爆发，都证明那些做出的假设过于简单的模型存在局限性。当然，我并非存心暗示你应该永远倾向复杂的模型，不该用那些简单的模型。正如本书前几章所讲，复杂的模型也会使人误入歧途。因为复杂的模型常常提供更明确（却未必更准确）的答案，这些答案会导致预测者过于相信自己，误以为自己可以预测得更准确，可实际上并非如此。

然而，简单化仍可算作一个模型的优点，但一个模型至少应该简单得高明些。SIR 这类模型尽管有助于了解疾病，但对于预测疾病的过程或许就无能为力了。

所有的预测都失败了

模型越复杂，就越能大幅提高预测的准确性，能证明这一点的例子相当少，气象预报便是其中之一。气象学家花费数十年的时间，创建了一个符合自然规律的物理模拟系统，有了这个系统，进行气象预测时，这些气象学家就能比单纯使用统计方法表现得更好。

现在，有越来越多的团队正试图用相似的方法预测疾病，他们采用的技术就是广为人知的“基于 Agent 建模方法”。我拜访了匹兹堡大学的一些研究人员，他们的研究一直处在这类技术开发的前沿。匹兹堡大学的团队将他们的模型命名为 FRED（即 framework for the reconstruction of epidemic dynamics 的首字母缩写，指传染病动态重构框架）。这个名字也是为了向电视节目《罗杰斯先生的邻居》的节目主持人、匹兹堡人佛莱德·罗杰斯致敬。

和芝加哥一样，匹兹堡也是一个社区之城。匹兹堡的研究人员一说到疾病，就会想到居民区，所以 FERD 模型本质上就是一个模拟的匹兹堡城——一项十分细致的模拟，每个人都用一个“仿真人”来代表，这个“仿真人”拥有自己的家庭、社交网络和居住空间，拥有一套和他的社会经济地位相匹配的一系列信仰和行为。

约翰·格雷芬斯特博士是匹兹堡团队中众多科学家中的一位，他的人生的大部分岁月都生活在匹兹堡，讲话时还带着当地独特的口音。格雷芬斯特博士向我解释 FERD 是如何组织起来的：“这个模型里有学校、工作场所和医院，它们都被放置在正确的地理位置上。这个模型的设置相当复杂，研究团队安排儿童都去上学，却不一定是就近入学——有的学校面积小，有的学校则面积非常大。因此，你会了解到这种合成的模拟城市人口数量。”

格雷芬斯特博士和他和蔼可亲的同事尚恩·布朗博士向我展示了 FRED 模拟的结果，在模拟的匹兹堡市、模拟的华盛顿城或是模拟的费城中，用颜色标注的疾病浪潮正从一个区域泛滥到另一片区域。但是，FRED 也是一项严肃的工作，这些模型也没有优于真实城市的虚拟处理方法：一个城镇、一座城市乃至一个州的每个人都用“仿真人”来代表。一些基于 Agent 建模方法甚至尝试着模拟整个国家，甚至整个世界。同气象预报模型一样，这需要进行大量的计算，因此需要借用超级计算机。

这些模型还需要借助大量的数据。模型需要准确的人口统计资料，通过人口普查就可以获得相当准确的相关资料。另外，这些模型还需要解释人类行为，而这项任务则要难以预料得多。例如，一位 26 岁的拉

蒂纳（意大利地名）单亲妈妈接种疫苗的可能性到底有多大？你可以设计一份调查问卷来询问她——基于 Agent 建模方法都十分依赖调查问卷的数据。但是在被问到健康选择时，受访者常会撒谎（或是记不清自己的选择）：比如，受访者宣称的洗手次数比实际次数要多，避孕套的实际使用情况也是如此。

格雷芬斯特博士告诉我，有一条原则相当深入人心，那就是人们接受并参与那些有些麻烦但有益健康的做法的意愿，与他们对患病风险的感知紧密相关。我们模拟的匹兹堡人只有在她认为这次 H1N1 流感十分严重时，才会去接种疫苗；若她觉得流感的情况不严重，她是不会去接种疫苗的。可是，如果她的邻居或是她的孩子生病了，她对流感的感知会发生怎样的变化？如果当地的新闻中充斥着流感的报道，她的反应又会怎样？因此，基于 Agent 建模方法与疾病预测中自我实现和自我否定的特性息息相关，因为它们是动态变化的，同时也允许这些仿真个体的行为随时间变化，在处理这些问题时，这些模型用起来也许更加得心应手。

另外，可以看看多姆博士和他在芝加哥大学的团队，他们正在建立基于 Agent 的模型，用来研究一种名为 MRSA（即耐药葡萄球菌）的危险病毒的传播，这种病毒会使普通的创伤，如切伤、刮伤、擦伤等演变为威胁生命，甚至无法治愈的传染病。MRSA 病毒的结构十分复杂，有很多传播途径：可以通过拥抱这类非常随意的接触传播，可以通过暴露的伤口传播，也可以通过汗液或血液这样的体液接触传播。有时这种病毒还会停留在很多物体的表面，如案台或毛巾。另外，运动员的更衣室也是孕育 MRSA 病毒的温床，在那里运动员常共用运动器材；已经报道的 MRSA 疫情有很多病例就发生在橄榄球队中，从美国高中橄榄球队到美国职业橄榄球联盟都有发生。但许多携带 MRSA 病毒的人并没有发病，也没有表现出任何病症，这便使情况变得更加复杂。

在为 MRSA 病毒建模时，多姆和他的同事必须问自己一些问题，比方说：哪一类人被割伤后会用邦迪创可贴？不同的文化背景下，拥抱行为是否普遍存在？一个社区中有多少人坐过牢（监狱中潜藏着葡萄状病毒）？

对于这类问题，传统模型根本无法解答，而基于 Agent 建模至少使做出较为准确预测成为可能。但是，匹兹堡和芝加哥团队需要考虑的变量数量大、范围广——当你竭力对整个人群中每个人的行为进行模拟时，必然会是这样的情形。这两个团队的工作常常要借用认知心理学、行为经济学、人种学乃至人类学的知识：基于 Agent 建模用于研究 HIV 在社区中的传播，而这些社区就如同巴布亚新几内亚的丛林般各具特

色，又如同阿姆斯特丹的同性恋酒吧一样风格迥异。要了解这些社区，需要对当地的习俗和环境等知识了如指掌。

因此，基于 Agent 建模对研究者的要求颇高，奋斗在这一领域的团队通常都是多学科交互的全明星队，团队成员都是各自研究领域的精英。但由于缺少数据，这些精英常常劳而无功。“就连 H1N1 病毒，我们都很难收集到有关何人、何时、何地患病的具体地理数据。”格雷芬斯特哀叹，“可想而知，要想获得过去爆发的所有疾病的数据有多困难。”

与匹兹堡和芝加哥这两个团队进行交流时，他们的话有时会让我想到一些关于中国的故事，这些故事你和我都读过，讲的是中国某些购物中心极尽装饰的情况——罗马圆柱、室内过山车，威尼斯运河，无一不有——但购物中心却没有店铺营业，也没有人逛。芝加哥和匹兹堡的团队都已经得出了一些十分有用且可行的结论——例如，匹兹堡大学的格雷芬斯特博士想到，过于草率、频繁地让学校停课（以防止疾病传染），有可能会事与愿违；而芝加哥大学的团队猜测到，芝加哥内城的 MRSA 病例数量不正常，这个问题是由进出库克县监狱的人群引起的。但是，这些模型模拟的情况至少都在几年之后才有可能出现，他们正翘首期盼着眼下那些根本不存在的数据。

基于 Agent 建模——不像气象预报的模型可以依据每日的数据进行修正——也很难检验。重大疫情只是偶尔来袭。即使这些模型是对的，但由于成功的疾病预测存在自我否定的特性，这些模型也有可能会成为自身成功的受害者。假设该模型暗示某项特殊的干预措施——如让学校停课——可能会十分有效，而这项干预措施也确实奏效了，现实中该疾病的传播速度就会降低。但回头想想，这样的结果可能会使该模型之前的预测看似过于悲观了。

因此，匹兹堡和芝加哥团队一直在犹豫是否将他们的模型应用到具体的预测中。其他团队在 2009 年 H1N1 流感爆发前则没有这么谨慎，其中一些团队发布的相关预测十分糟糕，严重地低估了 H1N1 病毒的传播范围。

目前，这些团队的活动大多局限于多姆的同事奇普·马卡尔所说的“为见解建模”。也就是说，基于 Agent 建模可以帮助我们进行了解传染病的实验，但目前这些模型还不能帮助我们预测疾病的爆发。

预测是为了让损失最小化

美国最近两次流感恐慌证明了，天花乱坠的宣传是不可靠的。1976 年，只有迪克斯堡爆发了甲型 H1N1 疫情，其他地区基本没有疫情；当

时的美国总统福特推行的“全民接种疫苗计划”纯属过激反应。2009 年，相当数量的人感染了 H1N1 流感病毒，但是病亡的病例却很少。这两个例子说明，美国政府关于疫情严重程度的预测实在有些离谱。

但谁都不能保证，下一次流感疫情爆发时，会不会出现同样的预测错误。虽然人类逐渐适应禽流感，但禽流感病毒 H5N1 仍会使数千万人丧生。若一种流感病毒的传播速度和 2009 年 H1N1 病毒的传播速度一样快，致死率又和 1918 年西班牙流感的致死率一样高，就会使 140 万名美国人丧生。还有其他一些威胁来自非流感病毒，如传染性非典型肺炎（SARS）病毒，甚至来自天花，虽然早在 1977 年全球就已经消灭了天花，但这种疾病极有可能被恐怖分子制成生化武器，再次引入社会，谋杀数百万人。根据定义，最严重的流行病传播的速度极快：2009 年，H1N1 病毒只用了大约一周时间就从医学界毫无察觉的病毒发展成一种让数百万人丧生的病毒。

我为本章写作所采访的那些流行病专家强烈地意识到自身所用模型的局限性，这一点和其他领域的专家形成鲜明对比。哈佛大学的马克·利普思奇告诉我：“依据 3 个数据点作预测是很愚蠢的。”利普思奇所说的“3 个点”是指 1918 年、1957 年和 1968 年 3 次流感的爆发，“我们能做的就是为病情发展的不同状况作好准备”。

如果你无法做出准确预测，却假装自己可以确保预测准确，这样做通常会贻害无穷。我猜想流行病专家和医学界的其他专家深知这一点，因为他们必须遵从希波克拉底誓言：最重要的是，不能伤害别人。

医学界对统计模型的使用最为谨慎，医生总能恰如其分地扮演预测的角色。这并不是说经济学家或地震学家作预测时没有风险。但由于医学直接与生命和死亡挂钩，医生会更加谨慎小心。在医学领域，错误的模型会让人丧命。后果很严重。

另外，奇普·马卡尔关于“为见解建模”的概念还要进一步充实。本书的主导思想是，预测既是手段也是目标。比如，预测在验证假说时发挥核心作用，因此在所有科学领域都发挥核心作用。

正如统计学家乔治·E·P·博克斯所写：“所有的模型都是错误的，但是其中有些是有用的。”这句话的意思就是，所有模型都是这个世界的简化形式，因为这是必要的。正如另一位数学家所说：“一只猫最好的模型就是一只猫。”其他模型都会遗漏一些细节。这些细节是否关系重大，取决于我们试图解决的问题究竟是什么，还要看我们想要得到的答案有多精确。

在我们使用的工具中，统计模型并非唯一一个要求我们做出粗略估算的工具。例如，语言就是一种模型，一种我们用于相互沟通的近似

值。所有语言都包含一些单词，在别的语言中却没有直接对应的同源词，即使两个词都在尽量解释同一个事物，它们也不是同源词。技术领域有其特有的语言。

但是，博克斯写道，一些模型是有用的。据我观察，芝加哥和匹兹堡两支团队正在进行的基于 Agent 建模就十分有用。这些模型能断定不同族群对疫苗的态度，能推出疾病在一个城市中不同社区的传播情况，也能猜到人们对流感新闻的反应，这些本身就是重要的问题。

一个好的模型即使失败了也有价值。“我们应当假设无论作什么预测，通常都是错误的。”欧祖诺夫告诉我，“所以，通常作预测就是为了了解错误是如何形成的，出错时该怎么做，如何将损失最小化。”

关键是要记住，模型是帮助我们理解某一领域复杂性的工具，而不是取代整个领域。这一点不仅在作预测时十分重要。一些神经学家，如麻省理工学院的托马斯·波焦认为人脑处理信息的方式就是通过一系列近似值获得的。

这就是为什么在预测时，形成更好的自我认识，正确解读所收到的信号如此重要。本书的前半部分主要介绍这些近似值在哪些领域服务于我们，又在哪些领域使我们惨败。本书的后半部分则关于如何更好地提高近似值，每次只提高一点。

第八章
贝叶斯定理：只有正确的预测才能让我们更接近真相

美国篮球职业联赛（NBA）的“赌神”哈若拉波斯·鲍勃·乌尔加利斯住在美国洛杉矶好莱坞山上的一座富丽堂皇、极富现代感的豪宅里，房子外部由金属和玻璃构造，房后还有一个游泳池，这里美得就像戴维·霍克尼油画中的景象。每年从 11 月到来年 6 月，乌尔加利斯每晚都会观看 NBA，每次看 5 场，分别由 5 台三星纯平电视机同时播放（就连直播电视公司也没见识过这种场面）。需要暂作休整时，乌尔加利斯会跑到拉斯韦加斯“棕榈树广场度假酒店”，在自己的公寓房中小住几日。如果需要放个长假，他就会跑到非洲来个徒步旅行。即使年景比较差的时候，乌尔加利斯差不多也能赚上 100 万美元，年景好的时候，赚上三四百万美元也不成问题。

乌尔加利斯很享受这种高品质的生活，但他并不符合那种穿休闲西装、嘴上叼着雪茄的老派赌徒形象。乌尔加利斯在下注时，从不依靠线人提示、贿赂裁判或是其他小伎俩，他也没有什么所谓的“秘诀”，他使用的是计算机模拟手段，但也并不是完全依赖这个手段。

乌尔加利斯之所以成功，是因为他分析信息的方式十分有效。乌尔加利斯并没有一味地寻求不同的预测模型，而是将自己掌握的统计学知识和篮球知识结合起来，并理清这些数据之间的重要关联。

这需要付出艰辛的努力，有时甚至会作不少无用功。乌尔加利斯能有今天的成功，还因为他下注时胆大、心细。

乌尔加利斯在加拿大马尼托巴的温尼伯长大，这座城市位于美国明尼苏达州边界以北约 145 公里，那里的人们工作勤劳却常受霜冻之苦。乌尔加利斯的父亲曾经非常富有，在其事业巅峰期时拥有 300 万美元身家，却因嗜赌将家产挥霍一空。在乌尔加利斯 12 岁那年，他的父亲破产了。乌尔加利斯 16 岁时意识到，要想离开温尼伯这个鬼地方，就必须接受良好的教育，而且只能自己挣学费。所以，考入马尼托巴大学之后，乌尔加利斯想尽各种办法挣钱。夏天，他会到不列颠哥伦比亚省最北部从事爬树的工作，每爬一棵树，可以挣到 7 美分。在校期间，他会去机场做行李搬运工，为去往多伦多或明尼阿波利斯或更远地方的温尼伯人来来回回地运送行李。

最后，乌尔加利斯终于存够了钱，买下了他做兼职的那家机场行李

搬运公司的部分股份，不久之后，他又拥有了更多股份。1999 年，读大学四年级的乌尔加利斯已经攒下了约 8 万美元。

但是，在乌尔加利斯眼中，8 万美元仍然算不得什么大钱，他曾经目睹父亲赢赢输输、进进出出好几回的钱都是这个数目。作为马尼托巴大学哲学专业的毕业生，就业前景也不是那么乐观。正在乌尔加利斯苦苦寻求加快其人生进程的途径时，他意外地参与了一场赌局，这场赌局令他欲罢不能。

那年，洛杉矶湖人队聘用了以打破陈规闻名的新教练费尔·杰克逊，费尔·杰克逊曾带领芝加哥公牛队赢得 6 次 NBA 总冠军。湖人队拥有大量实力球员：明星中锋、身高 2.13 米的“大鲨鱼”沙奎尔·奥尼尔，正处在最佳的运动状态；21 岁的后卫科比·布莱恩特，4 年前刚刚高中毕业，凭借自己的努力，成了一名越来越耀眼的明星球员。这两员大将——“大鲨鱼”奥尼尔和“得分王”科比——一直是湖人队在 NBA 中取胜的法宝，再配上杰克逊这样出色的教练，可谓“将遇良才”，杰克逊能使两人发挥出最高水平，对于湖人队来说更是如虎添翼。

然而，这种普遍的看法在湖人队身上却意外地落空了。前一个赛季，即因劳资纠纷、球员纷纷罢赛而导致赛季“缩水”的 1998~1999 年赛季，湖人队一直状态不佳，球队三易主帅，比赛胜负比为 31∶19，在第二轮季后赛中更是连输圣安东尼奥马刺队 4 场，被淘汰出局。与此同时，科比和奥尼尔“水火不容”，奥尼尔毫不掩饰自己对科比的嫉妒之心。科比还不到 21 周岁的合法饮酒年纪，人气却直逼久经沙场的“大鲨鱼”，在洛杉矶的运动商店里，科比球衣的销量也超过了奥尼尔，两人因此关系紧张。此时的西部联盟十分强大，拥有像圣安东尼奥马刺队和波特兰开拓者队这样紧密团结且经验丰富的球队，相比之下，湖人队则过于弱小，根本无法与之对抗。

在第三场常规赛中，湖人队不敌波特兰开拓者队。奥尼尔当即在赛场发飙，中场时被驱逐出场，这似乎印证了那些电视评论员和向来语不惊人死不休的节目主持人最担心的事。就连来自湖人队家乡的《洛杉矶时报》也开始批评湖人队，认为它只能排在 NBA 最佳球队的第 7 位，还责怪来自拉斯韦加斯的裁判人员在赛季开始前曾做出相对乐观的估计——湖人队会以 4∶1 的总比分赢得 NBA 总冠军。

进入 1999~2000 年常规赛季才几周的时间，拉斯韦加斯的篮球赌客们就开始持怀疑态度，他们将对湖人队的胜算提高到 6.5∶1，对于那些敢于挑战传统观念的人来说，可能会欠下巨额赌债。乌尔加利斯从不相信传统观念，他认为这些观念明显很荒唐，而正是这些传统观念中的缺点成就了他的成功。报纸的专栏记者和篮球赌客把太多的精力放在那些

不起眼的数据样本上，而忽视了数据周围更大的环境和情势。

乌尔加利斯并不觉得湖人队打得有多么差劲。尽管赛程十分艰难，而且需要不断地适应新教练，还要应对科比因伤缺阵的困局——科比在季前赛扭伤了手腕，此后一直没有打比赛。尽管出现了各种状况，可是湖人队在 7 场比赛中，依然获得了 5 场胜利。因为球员罢工和教练更迭，1998~1999 年赛季的湖人队可谓支离破碎，媒体焦点总是集中于此，却忽略了 1997~1998 年赛季比较正常的情况下，湖人队的胜负场比数达到 61∶21。乌尔加利斯观看过很多场湖人队的比赛，他很欣赏杰克逊的执教方式，所以他把自己的全部积蓄 8 万美元都押在湖人队身上（只留下很少一部分用作学费和生活费），赌湖人队这个赛季一定会赢得 NBA 总冠军。如果乌尔加利斯赌赢了，就能挣到 50 万美元，如果赌输了，他还得回到机场去做那份两班倒的工作。

起初，乌尔加利斯的直觉非常敏锐。他投下赌注时常规赛还剩下 71 场（共 82 场），而湖人队到赛季结束时赢得了其中的 62 场比赛，包括 3 次连胜——一次连赢 19 场，一次连赢 16 场，还有一次连赢 11 场。最终，湖人队以 67 胜 15 负的成绩结束了该赛季的比赛，这是 NBA 历史上常规赛季的最佳胜率之一。但在季后赛里，湖人队赢的就没有这么轻松了。那时的西部联盟各支球队的打法迅猛强劲，即使因为常规赛的突出表现而赢得主场优势的湖人队，想要在季后赛取得 4 连胜也并非易事。

季后赛第一轮，湖人队侥幸战胜了勇猛的萨克拉门托国王队，两支球队一直打到第五场决胜赛（当时的季后赛采取的是首轮五局三胜制，2001 年之后首轮也改为七局四胜制）才分出胜负。随后的西部联盟半决赛中，湖人队对阵菲尼克斯太阳队还算轻松。但是，在之后的一轮比赛中打得十分艰苦。湖人队拖住了波特兰开拓者队，开拓者队由迈克尔·乔丹的前任助手、杰克逊以前的学生斯科蒂·皮蓬带队，还拥有打法成熟的全能球员。尽管开拓者队缺少湖人队那样的天才球员，但是他们慢节奏的打法和强硬的身体对抗常常使其他球队乱了阵脚，总的来说，开拓者队仍然是湖人队的强劲对手。

决赛中，湖人队轻取七局四胜制系列赛的第一场比赛，可之后的表现就如同过山车一般。第二场对战波特兰开拓者队的比赛在洛杉矶举行，湖人队的表现糟糕得让人难以置信，在第三节比赛中连丢 20 分，整场比赛的比分是 77∶106，这是本赛季湖人队败得最惨的一场比赛。

之后的两场比赛是在波特兰的玫瑰花园球馆进行的。第三场比赛中，在上半场落后 13 分的情况下，湖人队众志成城，上演逆转好戏，由科比·布莱恩特在最后两秒投入绝杀一球。第四场比赛湖人队继续着

超水平的发挥，罚球命中率不堪入目的奥尼尔竟然也投中了所有的 9 个罚球，追平了两队之间 11 分的差距。就这样，湖人队将总比分扩大到了 3∶1，用杰克逊的一句不太恰当的话说就是，开拓者队已经站在了“死亡之门”的门口。

但是，到了在洛杉矶斯台普斯体育馆进行的第五场比赛，湖人队却始终无法把球投进篮筐。整场比赛共投篮 79 次，只有 30 次命中，最后以 88∶96 的成绩告败。第六场比赛又回到波特兰，湖人队仍然找不到状态，结果开拓者队以 103∶93 的比分轻松获胜。这样，双方又打成了平局，决定胜负的第七场比赛将在洛杉矶进行。

对于一个篮球赌客来说，最谨慎的做法就是对冲投注。比如，假设乌尔加利斯一面下注 8 万美元赌湖人队最终获胜，另一面下注 20 万美元赌开拓者队会在 2∶3 落后的情况下赢得第七场比赛，这样就会锁定乌尔加利斯所得的利润：乌尔加利斯不仅可以从对冲投注中收回投下的 8 万美元赌资，还会得到 22 万美元的净利润；如果开拓者队在 2∶3 落后的情况下没有最终获胜，而是湖人队最终获胜，乌尔加利斯最初投下的 8 万美元赌资会损失掉，他还可能会损失掉对冲投注的 20 万美元，但是仍然可以得到在两支球队上同时投注所得的 32 万美元的净利润。这个数目虽然无法与 50 万美元的目标相比，却也相当可观了。

但是这里有一个小问题：乌尔加利斯根本就没有 20 万美元，也不认识拥有这么多钱的人，至少在他可以信任的人里没有这样的有钱人。他只是一个 23 岁的机场行李搬运工，住在温尼伯哥哥家的地下室。很明显，他下注的结局只有两个，要么湖人队赢，要么他破产。

比赛之初，乌尔加利斯的运气并不是太好。开拓者队对奥尼尔处处盯防，想把他逼到罚球线上，对于湖人队来说，奥尼尔的每次罚球都是一次冒险（因为奥尼尔的罚球命中率低），或者让他为了避免因犯规满 6 次离场而备受束缚。在第二节的上半时比赛中，这种作战策略确实奏效了，奥尼尔犯规 3 次而且一直未能得分。第三节比赛中，皮蓬带领的开拓者队继续此前的凶猛攻势，最终以一个 3 分球结束了该节比赛，领先湖人队 16 分，斯台普斯体育馆嘘声四起。

此时，乌尔加利斯的胜算很低。湖人队在第三节比赛只剩两分钟时还落后 16 分，几乎没有哪支球队处在这样的困境中还能扳回比赛，据计算，湖人队失败的概率大概为 15∶1。乌尔加利斯的赌注——离开温尼伯的车票——似乎只能泡汤了。

但是，第四节比赛一开始，开拓者队依靠硬碰硬的身体对抗打法开始出现弊端，队员因不断的撞击体力消耗很大，身心俱疲。而湖人队是在主场作战，心理学家认为，主场作战会使运动员在必要时产生额外的

兴奋感。湖人队是一支年轻的队伍，他们的活力源源不断。

开拓者队一下子陷入了得分荒，第四节比赛开始后的 6 分钟内一分未得，而湖人队此时则加快了进攻的脚步，将比分差距缩小到个位数，先是缩小到 5 分，之后又缩小到 3 分，接着布莱恩·肖的 3 分远投将比分扳平，最后，科比的两记罚球使湖人队的得分反超。尽管比赛最后几分钟开拓者队的投篮命中率有所提高，但为时已晚，湖人队的两位巨星科比和奥尼尔完成了一次连贯的空中接力，锁定了比赛的胜局。

两周后，湖人队迅速完胜印第安纳步行者队，赢得了“魔术师”约翰逊时代的第一个 NBA 总冠军，而机场行李搬运工乌尔加利斯也加入了准百万富翁的行列。

成功的赌客是如何思考问题的？

乌尔加利斯怎么会知道湖人队能够获胜呢？其实他并不知道。成功的“赌客”以及任何领域中成功的预测者，从来不会以稳赚不赔的心态、无懈可击的理论和极其准确的尺度去看待未来，这些都是失败者的幻想，是过度自信的弊病。成功的“赌客”会把未来看成零星的可能性，这些可能性会像股票行情一样随着新信息的出现而上下浮动。当赌客们能够确保自己的预测不会出差错时，他们才会下注。

比如，乌尔加利斯下注时，拉斯韦加斯赌客们的态度就已经暗示了湖人队只有 13%的胜算赢得 NBA 总冠军。乌尔加利斯并不认为湖人队获胜的概率是 100%，但是乌尔加利斯确信的是，湖人队获胜的概率不止 13%，他认为获胜概率约为 25%。如果乌尔加利斯的计算是正确的，那么他的赌注从理论上说会赢得 7 万美元的利润。

表 **8–1** 乌尔加利斯眼中的湖人队赌局

结果	概率	净利润
湖人队胜利	25%	+520 000 美元
湖人队失败	75%	–80 000 美元
预计获利	+70 000 美元	

然而，如果将未来定位为概率色调梯度中的灰色，那么“现在”就是黑色和白色了。乌尔加利斯有 25%的概率获得 52 万美元，扣除有 75%的概率会损失掉的 8 万美元，得出的 7 万美元就是乌尔加利斯所能获得的理论上的利润。长期看来，输赢是对等的：对于一个出色的预测者来说，过去和未来的情况都可以依据长期概率表现出来，因此两者之间更

加相似，却与当下的情况有所不同。但是，这是一个“开弓就没有回头箭”的赌局，乌尔加利斯需要极大的把握（他认为有 6 个不同的理由可以解释赌客们低估了湖人队的实力），还需要保持极为清醒的头脑，这样才能达成目标。

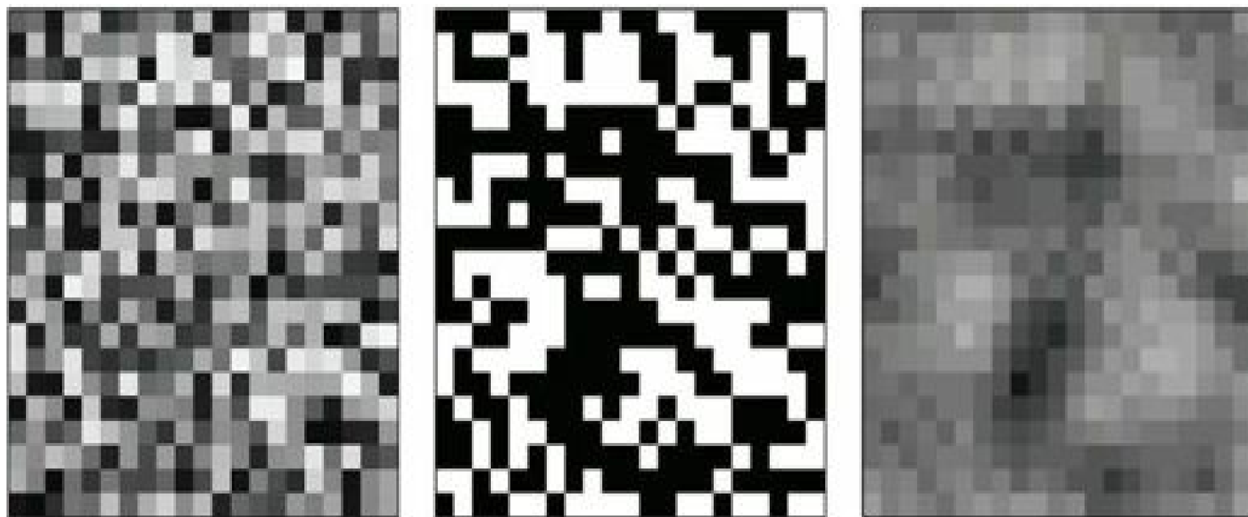


图 **8–1** 成功赌客眼中的世界

既然乌尔加利斯已经为自己筹集到了资金，他还是可以承担得起较小的风险的。乌尔加利斯可以在某个典型的 NBA 篮球之夜投三四个赌注。以任何正常的标准来看，这些都是极大的赌注，可与他获得的净利润相比，这些钱都是小钱，小到他看似根本不把这些钱放在眼里。我去拜访乌尔加利斯的那天晚上，他的眼睛聚精会神地盯着那些平板电视，眨都不眨一下，直到犹他爵士队将身高 2.19 米、动作不够灵活的乌克兰籍球员基里洛·费森科派上场，他才得空眨了眨眼睛，因为他知道，这一举动无疑表明爵士队已经认输了，而他自己也丢掉了押在这个队身上的 3 万美元。

乌尔加利斯的大秘诀在于，他根本就没有什么大秘诀，却有上千个小秘诀，他每次都把这些信息量子组合为一个矢量。比如，他有一个程序，这个程序可以对每一场比赛的结果进行模拟。但是，他并不完全依赖这个程序，只有在胜算很大或是有其他信息补充进来时，他才会依赖该程序。几乎所有 NBA 比赛他都看过——不论是直播还是录播——对于哪支队伍表现出色而哪支队伍表现不佳，他都有自己的看法。他经营着一家球探服务机构（本质上说其实是为他自己服务的），雇用了一些助手，让这些助手将每个球员在每场比赛中的防守阵形绘制成图，这种做法给乌尔加利斯带来的好处就连很多 NBA 球队都望尘莫及。他关注了几十名 NBA 球员的微博，仔细查看每一条 140 个字符的微博内容，试图从中找出关联信息：若某位球员发微博说自己那晚晚些时候会去某

个夜总会，那这位球员的心思很可能根本就没在比赛上。乌尔加利斯十分关注球队教练在新闻发布会上说的话及他们的措辞，比如，如果某位教练说，他希望他的队伍“学习进攻”或者“练好篮球基本功”，那可能表明他希望放慢比赛的节奏。

对于大多数人来说，乌尔加利斯观察的这些事物似乎十分琐碎。从某种意义上来说，那些大而明显的优势也会被其他篮球赌客发现，也会在盘分线上反映出来。所以，乌尔加利斯必须作进一步的研究。

比如，2002 年赛季末，乌尔加利斯注意到，凡是克里夫兰骑士队参加的比赛，比分都会超出下注的总分数（体育赌注主要有两种，一种是计算分数差，另一种是计算总分，即两支队伍得分的总和）。连着看了十几场比赛之后，乌尔加利斯很快就找到了原因：骑士队的得分后卫里基·戴维斯是出了名的自私自利，在赛季结束之后他即将成为自由球员，因此他所做的每件事都是为了提高自己的统计数据，尽可能地提升自己的价值。里基·戴维斯总是在千钧一发的时刻，努力组织进攻，制造各种机会增加自己的得分和助攻次数。比赛精彩与否已不再重要：骑士队已经丢掉了夺冠的机会。多数情况下，骑士队的对手也会丢掉夺冠机会，而且作为回报，双方会签订一份默契协议，在之后的比赛中，防守松懈、轮流投篮，从而提高自己的统计数据。该赛季最后 3 周里，骑士队参赛的比赛总得分突然从每场 192 分上升到每场 207 分。在总分上投注不是多么确切的事，当然，世上本就没有什么是确切的，但是总分投注却可以让投注人一本万利。

回顾过去，这样的比赛模式有时十分明显：如果骑士队除了不断地提高进攻数据以外什么都不做，那么，他们参加的比赛必会是高分比赛。篮球赌客在看统计数据时，观点极其偏执，从不考虑统计数据是在什么样的背景下产生的，于是这些球员就可以玩障眼法。如果一支球队一连串比赛的得分都很高，甚至出现三四次这种连续高分的情况，这样的高分通常没有任何意义。确实，因为 NBA 赛季很长，30 支球队要打 82 场常规赛，舞弊的现象时有发生。这些比赛大多为深水盘，这些情况的出现纯属偶然原因导致。而实际上，篮球赌客们通常也会认识到这些趋势，他们在划定盘分线时可能就会对这些趋势进行过度补偿，有时反方向下注才是明智的做法。

所以，乌尔加利斯不单单是寻找过去的比赛模式，在任何一个数据丰富的领域，寻找模式很容易，一般的赌客也都是这么做的。关键是要分辨出这些模式到底是噪声还是信号。

虽然还没有找到可以解释乌尔加利斯下注时作为依据的关键点，但是有一种思维过程可以帮助他作决定，这就是贝叶斯定理。

贝叶斯留下的宝贵遗产

托马斯·贝叶斯，大概于 1701 年出生（不过后人更倾向于 1702 年），他是一位英国牧师，也是英国皇家学会会员。虽然人们用贝叶斯的名字为概率论命名——贝叶斯定理或许是数理统计学中最著名的定理了——但对他的生平却知之甚少，甚至没有人知道贝叶斯的样貌，印在百科全书上的肖像大多也是张冠李戴，根本就不是他的真容。

贝叶斯出生在一个富足的家庭，他的家可能位于英国东南部的赫特福德郡，关于这一点人们基本上没有什么争议。因为贝叶斯不信奉英国国教，所以被牛津或剑桥这样的学府拒之门外，他只能不远万里到爱丁堡大学读书。

尽管贝叶斯所著图书的种类并不算多，但还是被选为英国皇家学会会员，在英国皇家学会，他担任内部评论家或者智力辩论的裁判员。尽管《神的慈爱》这篇短文是用约翰·努恩的署名发表的，但大部分学者认为这篇文章其实就是贝叶斯的作品。文中，贝叶斯思考了古老的神学问题：如果上帝真的是慈爱的，这世上为何还会有苦难和邪恶？贝叶斯给出的答案大体上是，我们不能将人类的瑕疵误认作上帝的缺陷，我们可能并不完全理解上帝所创造的这个世界。贝叶斯在给另一个神学家的回信中写道："所以一切看起来都那么奇怪……因为上帝只看到世界最底层的生活，他应该由此推断出整个人类会丧失幸福感。"

贝叶斯的作品《机会的学说概论》（又称《论有关机遇问题的求解》）的名气更大一些，直到 1763 贝叶斯去世之后，这部作品由他的朋友理查德·普莱斯引介到英国皇家学会，引起了学会的注意，随后才得以出版。这部作品主要研究的是，当我们遇到新数据时，该如何使用概率的方法进行推理。

在向众人呈现贝叶斯这一著作时，普莱斯举了一个"人"的例子，他是第一个出现在这个世界上的人（他可能是亚当，也可能是来自柏拉图洞穴的人），也是第一个看见日出的人。起初，这个人并不知道日出是必然现象还是偶然现象。然而，此后他度过的每一天太阳都会升起，于是他信心大增，认为这就是大自然的一个永恒特征。渐渐地，通过这一纯粹的统计学形式的推断，他预测太阳每天升起的概率为 100%（尽管从未达到 100%）。

贝叶斯和普莱斯并不认为这个世界本质上是盖然性的或不确定的。贝叶斯相信神是完美的，但他同时也支持牛顿的学说，认为大自然遵循一种有规律且可预测的法则。贝叶斯的理论更像是一种声明，从数学方面和哲学方面表达了我们是如何了解宇宙的：我们通过近似值一点点地

模拟并认识宇宙，收集越多的证据，就越接近真相。

这与苏格兰哲学家戴维·休谟的无神论观点形成了鲜明对比。休谟认为，既然我们不能确定太阳能否再次升起，那么认为太阳会升起和认为太阳不会升起的预测都是不合理的。贝叶斯则认为，合理性也是一种盖然论。其实贝叶斯和普莱斯是在告诉休谟，不要指责大自然，我们无法理解大自然只是因为我们不够聪明；如果你从无神论的牛角尖中走出来，并且对自然的规律作一些预测，也许就会离真相更近一些。

概率、预测与科学进步

我们可能注意到了，无神论的主张与贝叶斯在《神的慈爱》中提到的观点（我们不能把自己的不足怪罪到上帝头上）如出一辙。承认自我的缺陷才能补救不足。

然而，贝叶斯的哲学思想本质上没有任何宗教色彩。今天公认的贝叶斯定理就是一个普通的不能再普通的数学表达式，是由法国数学家、天文学家拉普拉斯推导出来的，而拉普拉斯就是一个无神论者。

也许你还记得，本书前文中提到拉普拉斯是科学决定论的倡导者。他认为只要给定宇宙内每一粒子的位置，并能快速计算其运动规律，人们就能完美地做出对宇宙的预测。那么，拉普拉斯为什么也与基于盖然论的理论撇不清关系呢？

原因在于，大自然完美无缺，人类对自然的认知和了解却缺憾万千，二者之间又断了联系。当天文观测显示木星和土星的运动轨迹出现异常时，拉普拉斯几乎要崩溃了，因为按照他的预测，木星即将撞上太阳，而土星即将飞入外太空。当然，这些预测都是错的，而拉普拉斯毕生都致力于测量这些行星的运行轨迹，力求得到更加准确的数据。那时，望远镜这样的仪器还不够精密，因此拉普拉斯所取得的进步只能依赖概率推理而非精确的测量。在拉普拉斯的眼中，概率是介于无知与博闻之间的基准点，更透彻地理解概率对科学进步极为重要。

18 世纪时，贝叶斯和拉普拉斯对概率、预测和科学进步之间的内在联系理解得更加透彻，借助几百年前发明的印刷机，人类社会由此开始进入信息大爆炸时期，并最终将这些信息应用于推动科学、技术和经济的持续进步。这种内在联系至关重要，与预测行星运行轨迹和湖人队能否夺冠一样重要。就像我们即将看到的那样，当另外一种不同的统计模型成为 20 世纪的主导时，这种模型不再强调预测的作用，而是试图将不确定性说成我们测量失误的结果，与我们错误的判断无关，科学发展就会受到牵绊。

简单的运算推导出重大的预测

如果说贝叶斯定理的哲学基础惊人地深厚，那么相比而言，其数学运算就少得可怜了。在其最基本的形式中，数学运算只是个代数表达式，包含 3 个已知变量和一个未知变量。然而，就是这样一个简单的运算，却可以推导出重大的预测。

贝叶斯定理涉及条件概率，也就是说，一旦发生了某个事件，这一定理就可以告诉我们一种理论或假设是否正确。

假设你和伴侣同住，某天出差回家后发现自己的衣橱里多出一件陌生的内衣。你可能会奇怪：自己的伴侣是不是出轨了？前提条件是，你找到了内衣，你想要评估的是自己的伴侣出轨的可能性。不论你相信与否，对于这样的问题，贝叶斯定理总能给出答案——假如你知道（或者有意愿预估）下列 3 个量：

> 第一，你需要预测出自己的伴侣在出轨的情况下，这件内衣出现的概率。为了解决这个问题，我们暂且假设你是一位女性，而你的伴侣是一位男性，那么，此时我们所说的内衣就是一件女式内衣。如果你的伴侣出轨了，那么很容易想象这件内衣是如何进入你的衣橱的。那么，即使他确实要做对不起你的事，你也希望他能够小心行事。在他确实背叛了你的情况下，我们认为，这件内衣出现的概率是 50%。
>
> 第二，你需要预测出自己的伴侣在没有出轨的情况下，这件内衣出现的概率。如果他没有出轨，有什么理由证明那件内衣的清白呢？当然有些理由会令人不快（比如这件内衣也有可能是他自己的）。或许，他把衣服搞混了；或者你的伴侣有一位红颜知己，两人之间只存在纯友谊，而你对此也深信不疑，她寄宿一晚忘了带走内衣；或者这就是你的伴侣给你准备的一件礼物，只不过忘了把它包起来。尽管这些理由有些荒谬，但也能说得通。你将这种情况出现的概率定为 5%。
>
> 第三，这点最为重要，你需要预测贝叶斯定理中所说的先验概率（或者简称先验）。在发现内衣之前，你认为自己的伴侣出轨的概率有多大？当然，现在很难完全客观地考虑这个问题，因为你已经发现了内衣。（在理想状态下，在开始查验证据之前，你就已经算出了先验概率。）但有时我们可以依据经验推断某事件发生的概率。比如，研究发现，已婚夫妇任何一年的出轨概率都在 4%左右，所以，我们可以将这个概率视为先验概率。

如果我们算出了以上 3 个概率值，就可以依据贝叶斯定理得出后验概率。令我们感兴趣的是这样的数据：在发现内衣的情况下，伴侣背叛我们的概率有多大？计算结果（和计算所得出的简单代数表达式）见表 8–2。

从图中可以看到，这一概率非常低：只有 29%，这个结果也许看似仍有悖常理——那件内衣果真是清白的吗？但这一概率之所以较低，是因为你把伴侣出轨的先验概率设定得很低。尽管一个清白的男人不能像出过轨的男人那样，能为一件陌生内衣的出现找出很多看似合理的解释，但你一开始就把他当作清白的人，这一点对方程式中影响很大。

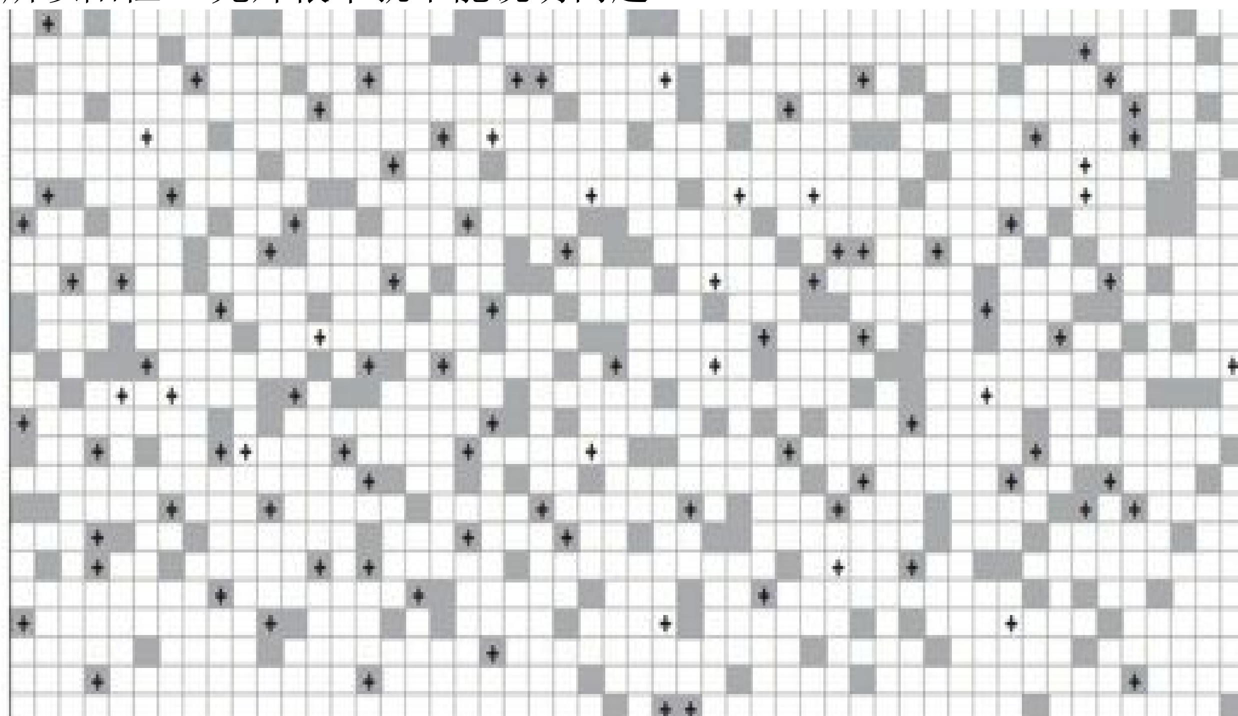
表 **8–2** 贝叶斯定理——内衣例子

先验概率		
男友出轨的初始概率预测	x	4%
新事件：发现神秘内衣		
在男友出轨的情况下，内衣出现的概率	y	50%
在男友未出轨的情况下，内衣出现的概率	z	5%
后验概率		
在你发现内衣的情况下，修正对男友出轨的预测值	$\frac{xy}{xy + z(1 - x)}$	29%

当我们的先验观念很强大时，它们在新出现的证据面前会表现出惊人的弹性。有一个经典的例子可以证明这一点，那就是女性 40 多岁时患上乳腺癌的概率。女性步入 40 岁之后，患乳腺癌的概率其实很低，只有 1.4%，这是很幸运的。但是，如果一位女性的乳房 X 光片显示阳性，那么她患乳腺癌的概率会是多少呢？

研究显示，如果一位女性未患乳腺癌，其乳房 X 光片会错误地显示她患乳腺癌的概率仅为 10%。而如果一位女性确实患有乳腺癌，X 光片会查出她患乳腺癌的概率约为 75%。看到这些统计结果，你会觉得阳性 X 光片似乎确实不是什么好消息，但如果用贝叶斯定理来分析这些数据，你会得到不同的结论：40 多岁的女性，即使乳房 X 光片呈阳性，其患乳腺癌的概率也只有 10%，因为鲜有女性年轻时就得上乳腺癌。因此，许多医生都建议女性在 50 多岁时再进行常规的乳房 X 光检查，而这样会使得患乳腺癌的先验概率更高。

这样的问题无疑极具挑战性。最近一项针对美国人的统计学意识的民意调查就介绍了这个乳腺癌的例子，结果发现，只有 3%的受访者能够给出正确的概率估值。有时，放慢速度直观地审视这个问题（如图 8–2 所示），反而会得到与不准确的估值完全相反的真实数值。可视化技术使得人们更容易考虑全局，因为乳腺癌在年轻女性中病发率极低，所以阳性 X 光片根本就不能说明问题。



正确假设（100/1 000）

- 肯定的统计学检验（正确的判断）(80/100)
- 否定的统计学检验（错误的判断）(20/100)

错误假设（900/1 000）

- 肯定的统计学检验（错误的判断）(180/900)
- 否定的统计学检验（正确的判断）(720/900)

图 **8–2** 贝叶斯定理**——**乳房 **X** 光片的例子

然而，我们通常会把焦点集中到最新、最快获得的信息上，而忽略了全局。鲍勃·乌尔加利斯这样聪明的赌客善于利用我们的这种思维缺陷。乌尔加利斯在湖人队的比赛上赌赢了，一部分原因是其他赌客过于关注湖人队最初的几场比赛。虽然有一员大将受伤，湖人队的表现仍然与你预想的强队应有的表现相差无几，可赌客还是将湖人队获胜的概率从 1/4 降到 1/6.5。贝叶斯定理要求我们认真考虑这些问题，当我们的原始近似值过于粗糙时，这一点很有用处。

不过，这并不能说明我们的先验概率总是支配新的证据，也不能说明贝叶斯定理本身会产生有悖常理的结果。有时，新证据的力量十分强大，会压倒所有其他证据，我们对一件事情的概率估计几乎可以立即从零跃升到 100%。

这里，我还要提到一个比较沉重的例子："9·11"恐怖袭击事件。2001 年 9 月 11 日清晨，当我们从梦中醒来时，大部分人都想不到恐怖分子的飞机会撞向曼哈顿世贸中心大楼。但是，世贸中心第一次遭遇袭击之后，我们才意识到这也许是一次恐怖袭击。直到第二座高楼被袭击之后，我们才相信确实遭遇了恐怖袭击。

贝叶斯定理可以复制这个结果。比如，在第一架飞机撞击大楼之前，我们预测曼哈顿的高楼遭遇恐怖袭击的概率只有 1∶20 000 或 0.005%。当然，我们还是会认为世贸中心意外遭遇飞机撞击的概率是非常低的。人们靠经验也能准确地预测出 0.005%这个数字：9 月 11 日之前的 25 000 天，一直有飞机盘旋在曼哈顿的上空，而期间只发生了两次这样的意外事故：一次是 1945 年的美国帝国大厦事件，另一次是 1946 年的川普大厦事件。这样看来，此类意外事故的日发生概率只有 1∶12 500。在第一架飞机撞上世贸中心大楼的那一刻，如果用贝叶斯定理计算这些数据（表 8–3A），发生恐怖袭击的概率便会从 0.005%剧增至 38%。

表 **8–3A** 贝叶斯定理**——**遭受恐怖袭击的例子

先验概率		
恐怖分子驾机撞击曼哈顿世贸中心大楼的初始概率预估	x	0.005%
新事件：第一架飞机撞击世贸中心大楼		
恐怖分子驾机袭击曼哈顿世贸中心大楼的概率	y	100%
恐怖分子未驾机袭击曼哈顿世贸中心大楼的概率（意外事故）	z	0.008%
后验概率		
在第一架飞机袭击世贸中心大楼的情况下，恐怖分子袭击曼哈顿世贸中心大楼的概率预估	$\frac{xy}{xy+z(1-x)}$	38%

然而，贝叶斯定理暗含的意思并不是说，我们对概率的预测只可以作一次更新，相反的，鉴于新证据的不断涌现，我们需要不断地更新自

己的预测结果。于是，第一次恐怖袭击的后验概率 38%，在第二次袭击之前就会变成先验概率。这时再来进行世贸中心遭遇第二次恐怖袭击的概率运算，我们遭遇袭击的概率就变成了 99.999%，这就表示恐怖袭击必会出现。在阳光灿烂的纽约出现意外事故的概率很低，而就像我们推断出来的可怕结果一样，第二次恐怖袭击很有可能会发生。

表 **8–3** 贝叶斯定理——遭受恐怖袭击的例子

先验概率		
在第一架飞机袭击世贸中心大楼的情况下，恐怖分子再次袭击曼哈顿世贸中心大楼的概率	x	38 %
新事件：第一架飞机撞击世贸中心大楼		
恐怖分子驾机袭击曼哈顿世贸中心大楼的概率	y	100 %
恐怖分子未驾机袭击曼哈顿世贸中心大楼的概率（意外事故）	z	0.008 %
后验概率		
在第二架飞机袭击世贸中心大楼的情况下，恐怖分子第三次袭击世贸中心大楼的概率	x y xy ＋ z（1 － x）	99.99 %

恐怖袭击、癌症、出轨等，这些富有挑战性的例子都是我精心挑选的，因为它们更能体现出贝叶斯定理的应用十分广泛。贝叶斯定理不是什么神奇的公式，在本书使用的简单形式中，无非是加、减、乘、除这些运算。我们还需要添加更多信息，特别是对先验概率的估计值，这样才能得出有用的结果。

即使是涉及我们不愿称为“偶然事件”的事件，贝叶斯定理也会要求我们用概率的方法思考问题。拉普拉斯认为，世间万物，不论是行星的运行轨迹，还是最小的分子运动，都是受牛顿定律支配的，这对于发展贝叶斯定理也是极有帮助的。不同的是，贝叶斯定理并不是要求我们认为世界在本质上和理论上都是不确定的，这一定理研究的是认识论的不确定性，也就是我们认识的局限性。

为什么大数据时代的预测更容易失败？

如果不能按照贝叶斯定理来思考问题，不单是乳房 X 光片会出现“假阳性”报告，所有科学都会出问题。2005 年，埃尼迪斯发表了一篇非常

有影响力的文章，题为“为什么大多数发表的研究成果都是骗人的”。埃尼迪斯在文中引用了大量统计论据和理论论据，就是为了说明医学期刊和其他学术或科学领域中，大量被视为真实的假设实际上都是不真实的。

正如我们提到的那样，埃尼迪斯的假设看上去还算是真实的。拜耳实验室发现，当他们试图利用实验再现医学期刊中的阳性结果时，却发现约 2/3 的结果都无法复制。检查一项研究发现是否真实的另一条途径是，看其在真实世界中能否做出准确的预测，正如本书所示，大多数情况下，这些发现都无法做出准确的预测。各个领域，从地震学到政治科学，预测的失败率实际上相当高。

埃尼迪斯告诉我：“过去 20 年里，可供使用的信息、基因组学和其他技术皆呈指数增长，有几百万个有趣的变量供我们测量。因此，我们希望利用这些信息和技术使预测成真，我这样讲并不是说我们过去没有取得多少进步，几百万份论文铺天盖地，如果真是没有什么进步，那将多么令人惭愧。但我们的新发现的数量明显比不上论文的数量，在创造新知识方面，大部分论文所做出的贡献真是微不足道。”

这也是为什么我们的预测在大数据时代更容易失败。拥有的信息量呈指数增长，需要验证的假设也正在以同样的速度增长。比如，美国政府现在发布了约 45 000 份关于经济的统计数据，如果你想要探究这些统计中所有两两组合之间的关系，比如亚拉巴马州的银行优惠贷款利率和失业率之间是否存在因果关系，则需要对 10 亿个假设进行验证。

但是，数据中那些有意义的关系组合——这里指的是因果关系而非相关性组合，而且这些组合能够证实这个世界是如何运转的——少之又少，增长的速度也不及信息本身的增长速度快，如今的真实信息也并不比互联网和印刷机问世之前多多少。大多数数据都只是噪声，就像宇宙的大部分都是真空区一样。

与此同时，就像贝叶斯定理所讲的那样，在某一个群体中，当某事的潜在发生率很低时（如年轻女性患乳腺癌的概率或庞大数据的真实性），如果我们不够小心，错误的判断就会主导事件的结果。图 8–3 生动地体现了这一点。在图中，80%的“真实”科学假设都被视为正确，而 90%的错误假设则遭到抵制，这看似没问题。然而，因为真实的发现非常少有，而其中却大约有 2/3 的发现被认为是真实的发现，其实是错误的。

不幸的是，就像埃尼迪斯指出的那样，大部分进行统计学检验的领域所发表的研究著作，大概都如图 8–3 所示。错误率为什么这么高？本书在一定程度上解答了这个问题。原因有很多，有些与我们的心理偏见

有关，有些与普遍的错误方法有关，还有一些与错误的动机有关。然而，归根结底是因为这些研究应用的统计学思维方式存在缺陷。

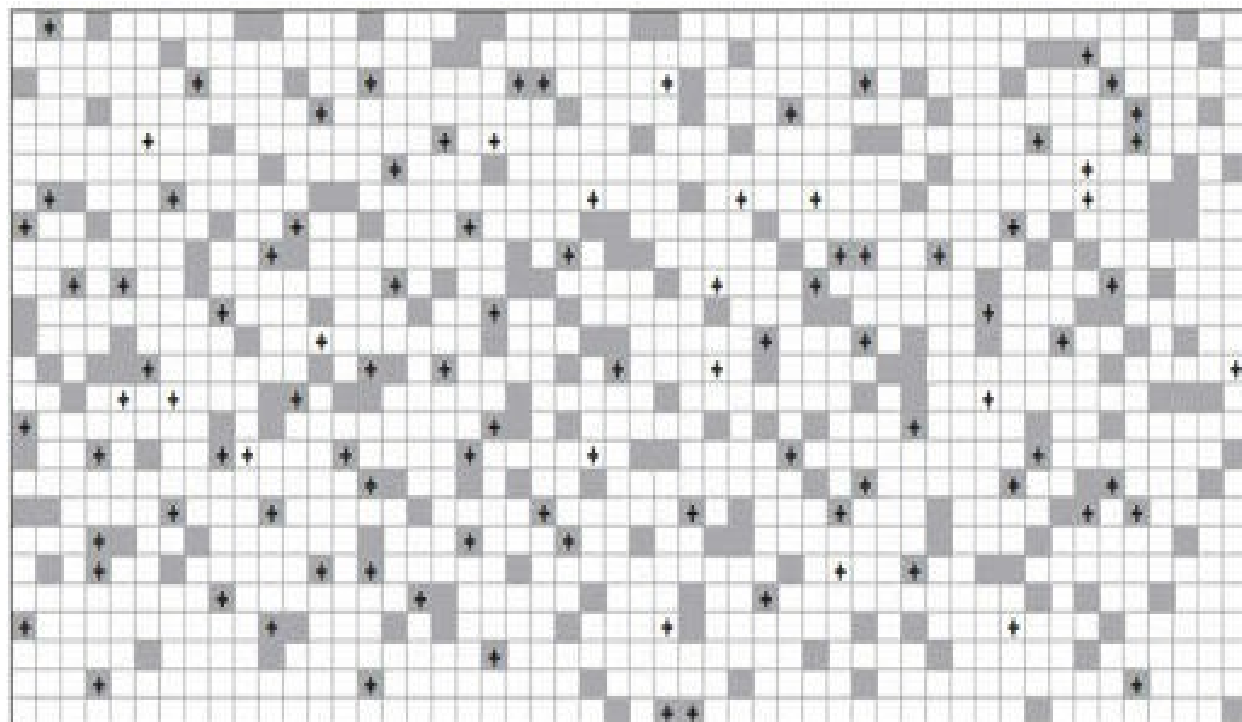


图 **8–3** 错误的判断图解

当统计数据偏离了贝叶斯定理

托马斯·贝叶斯最主要的思想劲敌大概要数英国统计学家及生物学家罗纳德·艾尔默·费希尔了。费希尔在贝叶斯去世将近 120 年后（1890 年）才出生，他天性活泼，几乎可以成为克里斯托弗·希钦斯笔下的英国传统智慧型人物。费希尔长相清秀却衣衫不整，不是叼着烟斗，就是抽着香烟，还时不时地与真实存在的竞争对手或假想敌发生冲突。他在讲课方面平淡无奇，可做起文章来却深刻透彻，在戏剧创作上也极具天赋，另外大家都很喜欢和他共同进餐。费希尔兴趣广泛，是当时最杰出的生物学家和遗传学家之一，但作为精英人物，他却毫不掩饰地抱怨社会贫困阶层的生育率高于知识分子阶层的生育率。（而他本人就有 8 个

孩子。）

统计学方法在当今之所以能够得到广泛使用，费希尔功不可没。他提出了“统计学显著性检测”的术语及方法论。虽然费希尔在他发表的一篇论文中第一次使用了“贝叶斯定理”这个术语，但其本意是想贬损贝叶斯。费希尔的另外一个主张是“将贝叶斯定理完全摒弃”，也可以说，他对贝叶斯和拉普拉斯完全不感兴趣。

费希尔和他的同代人在本质上对所谓的贝叶斯定理没有异议，因为贝叶斯定理不过是一个简单的数学公式。可贝叶斯定理的应用却让他们十分担忧，对贝叶斯先验概率这一概念格外忧心，他们认为这一概念似乎过于主观。

于是，费希尔等人力图建立一套统计学方法，为的是让我们不再受到主观偏见的干扰。现在，尽管这种统计学方法偶尔才会用到，但人们通常称之为“频率主义”。

“频率主义”的隐含意思就是，仅从人口样本（而非所有人口）中收集数据是导致统计学问题中出现不确定性的原因。这一点在政治民调中体现得淋漓尽致。在加利福尼亚州，有 800 万人要为即将到来的选举投票，如果只选出其中的 800 人进行抽样调查，结果就会出现人们所说的抽样误差。你在政治民调中看到的误差幅度就是用来量化抽样误差的：从 800 万人的投票结果中取出 800 份样本，究竟会出现多大的误差。“频率主义”的统计学方法正是用来量化误差的。

然而，即使是在政治民调中，抽样误差也并不总能反映事情的全貌。2008 年，艾奥瓦州民主党举行了决策会议，而新罕布什尔州进行了美国民主党初选，短暂的间歇时间里，在新罕布什尔州约有 15 000 人接受了调查——对于这个面积比较小的州来说，这个数目实在算得上巨大了，从理论上来讲，这个受访者的数量足以将误差幅度控制在±0.8%。然而，实际的误差却达到了 8%左右：民调显示希拉里在这个州的选票数会输给奥巴马 8 个点，而实际上，希拉里却以领先 3 个点的优势获胜了。抽样误差——这个“频率主义”唯一可以直接解释的错误类型——在新罕布什尔州的政治民调中可能只是一个小问题。

同样的，一些民调公司总是显示出对某一党派的倾向：他们可能对 2 亿美国成年人进行调查，但仍然得不到正确的结果。早在 250 年前，贝叶斯就解决了这个问题。如果使用的工具本身就带有偏见，那么作了多少测量并不重要，因为你的目标定错了。

从本质上看，“频率主义”解决统计学问题的方法是极力摆脱使预测出错的最常见原因——人为错误。“频率主义”认为不确定性是实验本身所固有的特质，而非我们认识真实世界的能力中所固有的特质。“频率

主义”的方法还意味着，你收集的数据越多，所犯的错误最终就会越趋近于零：这是解决所有问题的充分必要条件。本书中提到的预测问题比较严重的领域中，有用信息都十分稀缺，而收集更多的信息确实非常有价值。然而，如果使用不当，“频率主义”也不一定是通往完美统计的星光大道。正如埃尼迪斯指出的那样，大数据时代似乎只会使研究文献中的假成果问题更加严重。

不论在理论上还是实践中，“频率主义”的方法都不是很客观，要依赖大量的假设。种种方法总是假定，某一个测量方法中潜在的不确定性遵循贝尔曲线或正态分布，这是正确的假设，却不适用于股票市场。“频率主义”方法要求对样本总体下定义，在政治民调中，这一点十分明确，但在其他许多实际应用中，样本总体却存在很大的随意性。从“9·11”恐怖袭击事件中，人们能得到什么样的样本总体呢？

然而，更大的问题是，在力求设计完美无瑕的统计程序时，不能受研究者个人偏见的干扰，这就使得“频率主义”的方法总是与真实世界相隔离。这样的方法让研究者忽略了假设的基本环境和合理性，而这正是计算贝叶斯先验概率所必需的。于是，你必然会看到，一些发表了的权威论文中出现了蟾蜍是如何预测地震的，或是塔吉特百货这类大商场如何引起了种族仇恨等话题，并通过“频率主义”测验，最后产生了“具有统计学意义”（但显然是十分荒谬）的研究成果。

费希尔在其职业生涯末期变得成熟、温和，甚至偶尔还会赞扬贝叶斯一番。在费希尔漫长的职业生涯中，他的一些方法（尽管不包括当今广泛使用的方法）确实在贝叶斯定理和“频率主义”方法之间做出了妥协和让步。然而，在其生命的最后几年，费希尔出现了极为严重的错误判断，这表明了费希尔研究方法的局限性。

在吸烟和患肺癌之间的关系问题上，费希尔出现了失误。20 世纪 50 年代，人们作了大量研究，有些研究运用标准统计学方法，有些则运用贝叶斯定理，这些研究都表明吸烟与患肺癌有关，这一点在当今社会已被广泛认同。

费希尔晚年的很大一部分时间都在跟这些结论作斗争，他在《英国医学会杂志》和《自然》等颇具声望的杂志上发表了许多文章，旨在证明这一结论有误。他并不否认吸烟与患肺癌之间的统计学关系十分显著，而正如英国的苹果进口量和结婚率之间的历史关联性一样，他认为这只是把相关性误认为是因果关系，他一度还认为，是肺癌让人们更想吸烟，而不是吸烟引起了肺癌。这一思想明显是让人们无忧无虑地继续吸烟，而忘记了肺癌的困扰。

许多现在被广泛接受的科学发现，曾一度被视为一派胡言。有时是

因为当时的文化禁忌（如伽利略的日心说），但多数时候是因为可用于分析问题的数据并不存在。如果直到 20 世纪 50 年代还没有明显的证据证明吸烟与肺癌之间的必然联系，那我们就不应该质疑费希尔。然而，一些学者重新调查了当时已有的迹象并得出结论，这两者之间确实存在必然联系——许多研究人员在许多领域进行了各种统计学测验及临床试验，得出的结论都表明吸烟与肺癌之间存在因果关系。这一个观点迅速成为科学共识。

那么，费希尔为什么拒不接受这一理论呢？或许他是烟草公司的受薪顾问？或许他本人就是终身烟民？或许费希尔喜欢与别人背道而驰，喜欢备受争议，而且不愿受清教主义的束缚？简而言之，费希尔在许多方面都心存偏见。

但更大的问题可能是费希尔的统计方法更倾向于假想世界，更注重实验的客观程度——只有收集到足够的数据，所有假设才能通过检验并得出完美的结论。然而，要想达到费希尔要求的那种客观程度，就不得不否认贝叶斯先验概率的必要性，还要与其他混乱的实际环境脱节。这样的方法既不要求也不鼓励我们考虑假设的合理性：认为吸烟引起肺癌与认为蟾蜍可以预测地震这两种观点相类似。费希尔意识到相关性并不总是代表因果关系，这一点是值得赞扬的，但是，费希尔的统计方法并没有鼓励我们认真地考虑哪些相关性是因果关系，哪些不是。费希尔倾其一生都以这样的统计学方法思考问题，也难怪他始终无法辨别相关性与因果关系的区别。

成功践行贝叶斯定理的体育赌客

在贝叶斯的世界观里，预测是衡量进步的标尺。我们可能无法得到百分之百的真相，但只有正确的预测才能让我们更加接近真相。

贝叶斯对赌客十分尊重。和其他早期的概率理论家一样，贝叶斯和拉普拉斯经常利用概率游戏的例子来解释自己的工作。（尽管贝叶斯本人也许不怎么爱打赌，但他更喜欢纸牌或桌球这类大众玩的赌钱游戏。）赌客作（准确的）预测，作（非常准确的）概率预测，当他决定按照自己（更准确）的预测下注时，他的世界观就暴露在所有人面前。对贝叶斯先验概率最实用的定义大概就是下注的胜率。

乌尔加利斯是典型的“贝叶斯”式赌客，他喜欢精确地对篮球比赛下注，因为他可以用这种方式进行自我检验，也可以检验其下注的准确度。在我们的谈话接近尾声时，乌尔加利斯说道：“如果你是某支球队的总经理，你可能会这样说，这个队员我要了，那个队员我也要了。一天下来，你也不知道这样做是否正确，但我并不是这样做的，每天或每

个赛季结束时，我都知道自己的决策是对还是错，因为我知道自己是在赢钱还是在输钱。这个方法相当灵验。”

乌尔加利斯总是想尽办法搜集篮球信息，因为任何事都可能改变他的概率估值。像乌尔加利斯这样的职业竞技体育赌客，只有在认为胜算达到 54%以上时才会下注，因为这样才足以抵消“抽头”（博彩经营者从赌客的赢利中抽的份子钱）和下注风险。凭借着高超的技艺和勤奋的工作，乌尔加利斯跻身当今世界最成功的竞技体育赌客之列，但其下注的准确率仅为 57%。要想超过这一数字，比登天还难。

只要能使乌尔加利斯将自己的胜算从 53%提高到 56%，哪怕是一条很不起眼的信息也会因此变得关系重大。不管是玩扑克牌还是玩股票，玩家就指望这点薄利维持生计了。费希尔所说的统计学意义，指的是使用任意截点来断定哪些才是“有意义的”研究结果，哪些没有意义并且缺乏研究背景，这对赌客们来说实在太复杂难懂了。

但这并不是说乌尔加利斯提出假设时，有意地避开了他在统计数据中看到的信息（费希尔有关假设检验的见解有问题，但问题不在于假设本身，而在于他推荐的检验方法）。实际上，这一点对乌尔加利斯的工作至关重要。每个人都能看到统计模型，这些模型很快会在盘分线上体现出来，问题是这些统计模型代表的是信号还是噪声。乌尔加利斯根据其篮球知识提出假设，所以他能更快、更准确地分辨信号和噪声。

从表 8–4 中可以看到，乌尔加利斯赌球的方法是一种科学方法，属于纯净馏的一种。首先观察世界，之后提出问题：为什么克利夫兰骑士队不断地赢得比赛？接着，针对这一问题搜集信息，提出假设：这是因为戴维斯正处在合约年，他极力想尽快提高自己的统计数据。乌尔加利斯和物理学家或生物学家的不同之处在于，他以下注的方式验证自己的预测，而科学家们更喜欢通过实验证实自己的预测。

表 **8–4** 科学方法

科学方法的步骤	竞技运动博彩案例
观察一种现象	骑士队不断赢得比赛
提出假设解释这一现象	骑士队不断赢得比赛是因为戴维斯即将签署新合同，所以想尽可能得高分
根据假设做出预测	直到赛季结束，戴维斯始终保持同样的状态。因此，1）他将继续以快节奏 打球，2）骑士队未来将继续打出高比分的比赛

如果乌尔加利斯愿意根据自己所看到的数据提出有力的假设，他就会更积极地下注。比如，假设乌尔加利斯从丹佛掘金队教练的口中读出一些信号，如将会为球迷们呈现一场“精彩绝伦的比赛”。这句话可能只是随口说说，但也有可能是在暗示，掘金队会加快比赛节奏，以拉动门票销售。如果这一假设是正确的，乌尔加利斯对掘金队的赌注胜算将达到 70%，而不是通常的 50%。因为贝叶斯定理的缘故，乌尔加利斯越信任其假设，从掘金队身上获利的速度就越快。他可能要看过一两场比赛之后才会下注，因为他要观察自己的理论在实践中是否成立，即使只使用这种方法，拉斯韦加斯的赌客们也已经望尘莫及了。相反的，乌尔加利斯要避免被统计模型干扰，正如 1999 年湖人队开场不利，这场比赛本无多少深层意义，但其他赌客却误把它当作了信号。

通往真理的贝叶斯之路

乌尔加利斯的概率预测是主观的还是客观的呢？这很难说。

按照经验来看，所有人都有各自的信仰和偏见，这种信仰和偏见是由个人的阅历、价值观、知识涵养、政治立场或专业背景等因素凝聚而成。贝叶斯定理明确地承认人们固有的观念会影响对新证据的理解，还生动地描述了人们对于世界的变化做出的反应，这是贝叶斯定理的一大优点。比如，假设费希尔认为吸烟引起肺癌的概率只有 0.000 01%，这就解释了为什么在所有证据都证明吸烟确实会引起肺癌时，这位先生仍然不信。实际上，贝叶斯定理并不干涉个人信仰，在这一理论的支持下，没有什么可以阻止你将自己认为绝对正确的事物作为自己的信仰。如果你认为上帝百分之百存在，或者上帝根本就不存在，在贝叶斯定理下，所有的证据都不会劝你改变观点。

我并非在这里告诉你对有些事物是否应当持有绝对明确的信任态度，但或许我们应该更加坦诚地说出想法。完全相信某件事的人和完全不相信某件事的人争论不休，根本就是无益之举，许多战争或许就是由这样的争论引起的（如印刷机问世之初发生在欧洲的“宗派战争”）。

这并不是说所有的先验理念都是正确无误或者行之有效的，但我认为，我们的信念使得我们永远做不到完全客观、合理或是准确。不过，我们可以尽量做到少一些主观、少一些不合理、少一些错误。基于个人信念之上的预测，是自我检验的最佳（或许也是唯一的）方法。如果客观性关注的是超越我们个人处境的更大真理，而预测是检验我们的认知是否符合真理的最佳方式，那么我们当中最为客观的人，就是作预测最准确的人。费希尔的统计学方法认为，客观性存在于实验室的实验当

中，所以与贝叶斯定理相比，费希尔的统计学方法并不适合完成准确预测这一任务。

实际上，贝叶斯定理有一个特性，即随着时间推移，证据越来越多，我们的观念也应该彼此交合。在图 8–4 中，我给出了一个例子，3 个投资者正在判断自己是处于牛市还是熊市。他们最初的看法截然不同——其中一个投资者非常乐观，他从一开始就认为股票市场有 90%的概率是牛市；还有一个投资者的脾气有些暴躁，认为股市只有 10%的概率是牛市。每次股市上涨，投资者的情绪就比之前乐观一点；每次股市下跌，他们的情绪就悲观一点。然而，在我设定的这套模拟机制中，尽管每天股市上下振荡，但总的来说还是上涨了 60%。虽然道路坎坷，但 3 位投资者最终还是正确地做出了判断，他们几乎（尽管并非完全）确信自己正处于牛市当中。

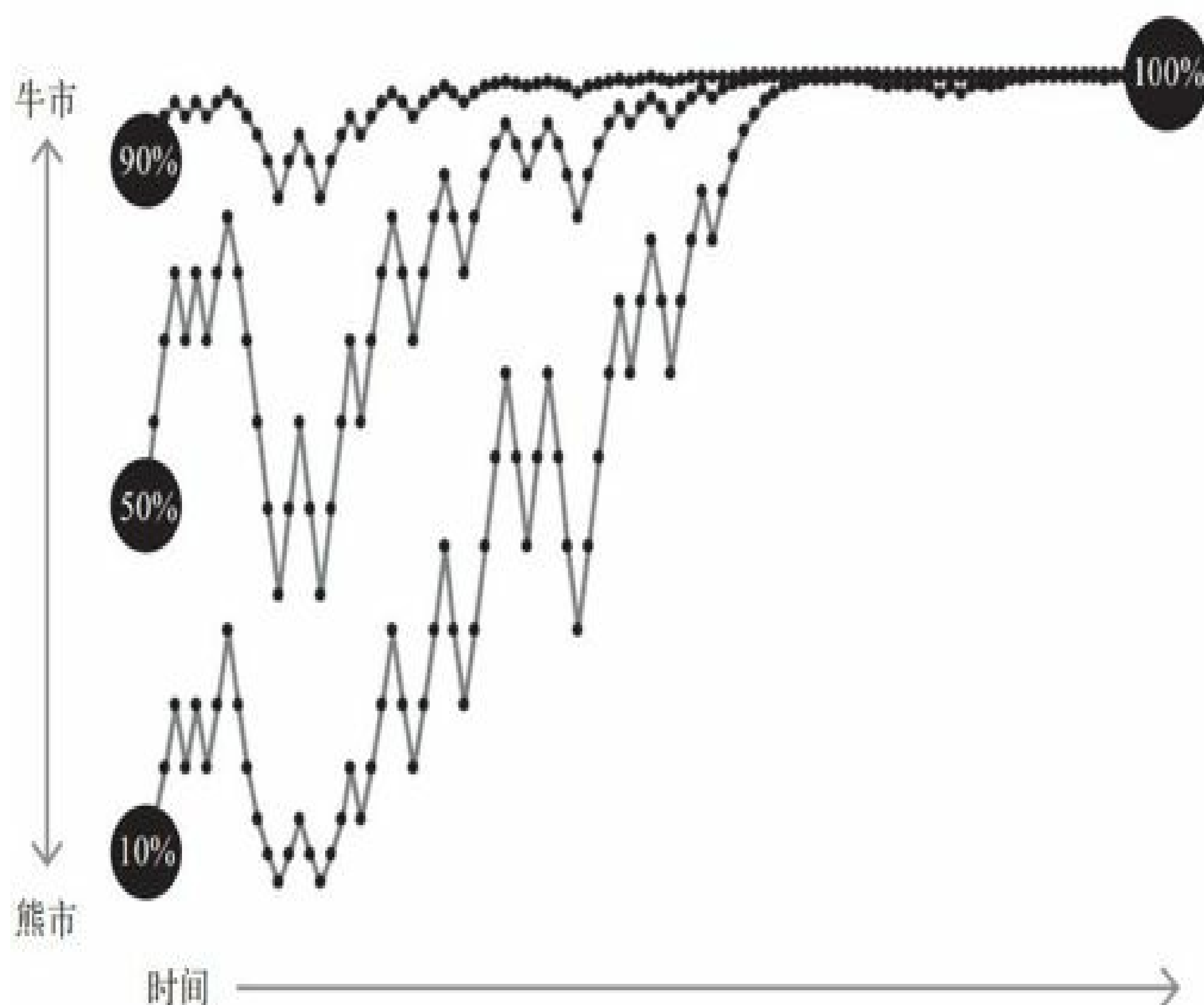


图 **8-4** 贝叶斯定理的收敛性

从理论上来讲，科学就应该这样产生作用。达成共识的科学概念难以琢磨，但真理越辩越明，新证据不断出现，科学观点总能汇聚到一起并且向真理逼近，一如股市的曲折反复。科学界在对范例进行改编、使之与新证据相符时，虽然有时也操之过急，但通常总是过于保守。而不变的是，只要踏上了贝叶斯之旅，所有正确的甚至是错误的固有观念都会重新得到审视，并最终走向真理。

举个例子来说，此时很多科学家正在使用的统计学方法或许正在进行范式转移。我在这里对费希尔统计学方法中出现的瑕疵所作的批评并不新奇，也不过激：多年来，不论是临床心理学、政治科学，还是生态学领域的著名学者，总在发出这样的声讨，但迄今为止统计学方法仍没有任何根本性的改进。

然而，最近一些德高望重的统计学家认为，“频率主义”统计学方法不再适用于大学课程。另有一些教育学者也在考虑将费希尔定理从学报中移除。事实上，近 10 年出版的各类刊物都在向人们宣传贝叶斯定理。

乌尔加利斯也是一样。虽然不是每次作预测都遵循贝叶斯定理，但是他在假设中检验统计数据，从篮球知识中获取认知方法，甚至乐于接受概率性答案，这些完全是贝叶斯定理倡导的做法。

教科书和传统观念不会说变就变，但贝叶斯定理始终认为，我们定会找到更好的方法，因为贝叶斯定理预测到贝叶斯理论家终会大获全胜。

第九章

国际象棋大战：计算机与人类的智能博弈

和很多前辈一样，24 岁的埃德加·艾伦·波也为亚马逊土耳其机器人着迷，这是一台会下国际象棋的新奇机器，拿破仑·波拿巴和本杰明·富兰克林都曾是它的手下败将。1770 年这台机器在匈牙利问世，那时艾伦·波还未出生，美国也未建立。19 世纪 30 年代，这台机器来到美国海港城市巴尔的摩和弗吉尼亚州首府里士满巡回展览，而此前几十年这种机器就已经风靡欧洲了。艾伦·波推断这不过是一个精巧的骗局，齿轮和传送装置暴露了亚马逊土耳其机器人的秘密——一定有一位象棋高手坐在箱子里，操纵杠杆，移动棋子，每次“将军”时，方巾包裹的头就会向下点一下。

艾伦·波被视为“侦探小说之父”，他的有些推理作品神秘而离奇。比方说，这次他的怀疑就是有据可依的。艾伦·波怀疑一个人（后来他断定此人便是德国国际象棋大师威廉·斯伦贝谢），在拆卸和组装机器时都有他的身影，但比赛一开始他就不见了踪影（啊哈！原来是威廉·斯伦贝谢在箱子里）。

然而，在艾伦·波有关亚马逊土耳其机器人的文章中，真正有预见性的观点，应该是他抓住了这个机械人对我们今天所说的人工智能（这一术语直到 120 年之后才出现）的深层影响。计算机未来有望能模拟甚至改进人的高级功能，艾伦·波的文章对这一前景表达了非常深刻、非常现代的矛盾心理。

艾伦·波承认，若真有一台机器能下国际象棋，那将十分引人注目。在他写那篇文章时，人们口中“与众不同”的第一台机械计算机——查尔斯·巴贝奇称为差分机的机器——都还未被构想出来。巴贝奇想过要制造计算机，但终其一生都未能完整地制造出来，而且，这台计算机除了完成加、减、乘、除的简单运算外，最多只能粗略估计一些基本函数的值，比如对数运算。艾伦·波认为巴贝奇的想法已经很吸引人了，但归根结底，这台机器能做的不过是输入可预测的信息，转动齿轮，然后再输出可预见到的结果。这毫无智慧可言，纯粹是机械计算而已。但是，若一台计算机会下国际象棋，就是一个奇迹，因为下棋需要智慧。

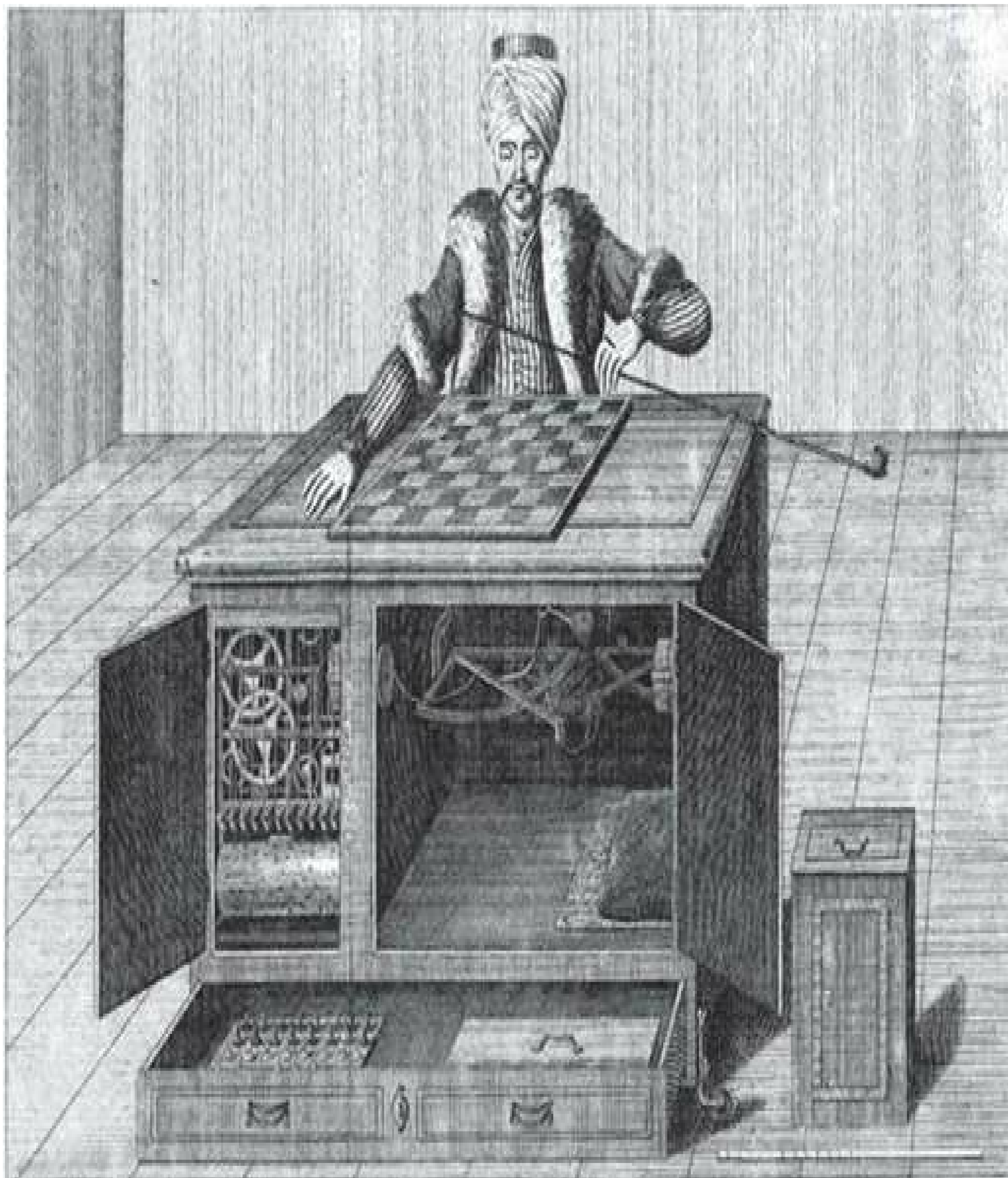


图 **9–1** 土耳其机器人

艾伦·波说道，如果真的有这样一台下棋机器，从定义上看，这台机器肯定局局获胜，因为机器是不会犯计算错误的。但是，这个亚马逊土耳其机器人并非招招完美，艾伦将这一事实当作进一步证伪的证据：这不是一台机器，而是由人控制的器械，充满人性的缺点。

尽管艾伦·波的逻辑也有缺陷，但时至今日人们仍对机器心存敬畏。

在所有表现人类智慧的一流发明中，计算机最震撼人心。比尔·盖茨常被票选为“全美最令人敬佩的人”，而苹果和谷歌则当选“最受欢迎的公司”。我们总是期待计算机既省力又完美，甚至能够克服人类的缺点。

另外，人们相信电脑程序的计算结果百分之百的准确，甚至是有预见性的。2012 年，两个英国青少年被指控行骗，他们向股票投资者推广选股机器人“MARL”,称它能“进行每秒 1 986 832 次的数学计算”，同时可以避免“人类的直觉判断”，投资者只要购买了 MARL 推荐的低价股票，每过 1 个小时投资就能翻倍。最终两个少年成功骗取了 100 多万美元。

即使计算机预测没有使人们更容易被骗，大家对它仍心怀敬畏。比如，一些计算机能预测医院病人康复的概率。在新闻报道中，这些计算机常被表述成现实版的人工智能哈儿 9 000（电影《2001：太空漫游》中的计算机）。在电影中，哈儿 9 000 认为自己对宇航员不再有什么用处了，于是便试图遏制他们的活动。

进入大数据时代，信息爆炸，人们处理信息的能力迅速提高。在这样的时代，我们对待计算机及其可能为人类做出的贡献也许应该抱有更加健康的心态。技术作为节省劳动力的手段，确实可以惠及人类，但我们不该期待机器能代替我们进行思考。

国际象棋的人机大战

1912 年，西班牙工程师莱昂纳多·托雷斯·克维多制造了另一个版本的土耳其机器人，将其命名为“棋手”。尽管常有人将棋手这个机器人视为第一款国际象棋计算机游戏，但是它的功能非常有限，只能处理剩下 3 个棋子的残局，决定这 3 个棋子的走步。“棋手”也没有像亚马逊土耳其机器人那样的标志性头饰。）

现代国际象棋计算机之父是来自麻省理工学院的数学家克劳德·香农，他也是信息论的创立者。1950 年，香农发表了一篇题为“设计程序让计算机下棋”的论文。香农发现了一些电脑程序方面的运算法则和技术，至今仍是国际象棋计算机游戏程序的核心。香农还明白了为什么国际象棋会成为检测机器处理信息能力的对象。

香农发现，国际象棋的目标出奇地清晰、明了，就是把对方的王“将死”。另外，国际象棋的比赛规则也相对简单，比赛中没有运气成分和随机因素。然而，只要下过国际象棋的人（我本人也只是略知一二）都会发现，国际象棋目标简单、规则简单，但想赢得比赛却并不是那么容易的。在一局棋中要走上 24 步都需要高度集中精神，更不要说赢得比赛了。香农将国际象棋视为计算机界的“试金石”，直接用它来检验计算

机的计算能力及其可能拥有的其他能力。

但是香农与其后继者不同，他并没有对计算机持有过于理想化的浪漫观点，香农并不认为计算机能像人类一样布阵下棋。同时，香农觉得计算机下棋赢了人类并非必然。相反，他认为计算机有以下 4 个优点：

1. 计算速度快。
2. 不会犯错，除非编程时就编入错误。
3. 不会偷懒，在分析招数、分析可能位置时不会半途而废。
4. 不带感情色彩，不会赢了一步就过度自信以致失去胜势，或是遇到困局就沮丧，劣势其实是可以逆转的。

香农认为，计算机的这些优点可以与人类具备的 4 大优势相抗衡：

1. 思维灵活，解决问题知道变通，不会按部就班。
2. 拥有想象力。
3. 懂推理。
4. 会学习。

香农认为，人机竞赛是公平的。但是在 20 世纪 90 年代，这种情况只是少数，且稍纵即逝，例如，俄罗斯国际象棋大师加里·卡斯帕罗夫是历史上最厉害的棋手，他曾对阵 IBM 公司的计算机“深蓝”，深蓝是史上最先进的计算机之一。

在这场比赛开始之前，人类一直保持着人机竞赛的不败纪录，甚至连平局的情况都没出现过。但是，从那场比赛之后，计算机却超越了人类，而且，在我们的有生之年会一直领先下去。

国际象棋比赛预测和启发法

根据贝叶斯定理，预测基本上属于信息处理活动——用新数据检测关于客观世界的假设，目的是为了更加真实、准确地理解世界。

国际象棋也许可以被视为一种与预测相类似的活动，国际象棋的棋手也必须处理信息——32 个棋子的位置和所有可能的走法，并利用这些信息制定出“将军”的策略。这些策略本质上就是假设——如何赢得棋局的假设。谁的假设更好，谁就能赢得比赛。

国际象棋是确定性的游戏，没有运气的成分。正如本书前文中所

写，从理论上讲，天气系统也是如此。可是，我们对这两大系统的了解远远不够。尽管已经十分熟悉支配天气系统的规则，但对构成云系、暴雨和飓风的每一个分子的位置，我们并没有完整的信息。因此，我们只能进行概率性预测。

在国际象棋中，因为棋子的数量有限，且位置显而易见，因此全部游戏规则和完整信息尽在我们的掌握之中。但是，国际象棋对技术的要求很高。这一运动项目证明人类处理信息的能力是有限的——可能也在告诉我们，尽管存在这些局限，我们还是可以制定出最佳决策。人类需要预测，未必是由于世界本身的不确定性，而是因为全面了解世界已经超出人类的能力范围。

于是，无论是电脑程序还是人类国际象棋大师都要依靠简化手段来预测比赛结果。这些简化手段可被视为“模型”，而在研究电脑程序设计和人类决策过程时，人们更愿意使用“启发法”这一术语。这个词源自希腊语中一个同根词，由此派生出另一个词 eureka（意为“我发现了”）。当一个问题的确定性解决方案超出了我们的实际能力时，我们就会采取依据经验法则的启发法解决问题。

启发法尽管十分有用，但必然会产生偏见和盲点。例如，“一旦碰见危险动物，立刻逃跑”。这是一条很有用的策略，可是，若你遇见的是北美洲灰熊，突然的移动就会惊吓到这个大家伙，它反而会跑过来追你。（相反，美国国家公园管理局提示游客，遇见灰熊时应尽可能地保持安静，必要时需躺下来装死。）在下棋时，人脑和电脑采用不同的启发法。人机对抗时，比赛的胜负就在于谁先发现对方的盲点。

从 1986 年到 2005 年退休之前，加里·卡斯帕罗夫一直是国际象棋界的顶级大师。1988 年 1 月，加里·卡斯帕罗夫预测出至少得等到 2000 年才会出现能战胜人类象棋大师的计算机程序。在巴黎召开的一次会议上，加里·卡斯帕罗夫带着嘲弄的语气说道：“如果哪位大师对战计算机时遭遇困难，我很乐意充当军师。”然而，就在 1988 年年底，丹麦的国际象棋大师本特·拉森被一个名为“深思”的计算机程序打败，这个程序是一项由研究生设计的项目成果，它是由卡内基梅隆大学的几个学生设计的。

然而，“深思”程序打败的不过是普通的人类国际象棋大师，而不是顶级大师卡斯帕罗夫，1989 年“深思”迎战卡斯帕罗夫时，输得一塌糊涂。卡斯帕罗夫向来尊敬计算技术在国际象棋中的作用，也一直向计算机学习以提高棋艺，但他对“深思”却少有赞誉，只是说希望有朝一日能出现一台需要他“用尽全力”才能战胜的计算机。

由许峰雄和默里·坎贝尔领衔的“深思”设计团队最终受雇于 IBM，在

那里他们将“深思”的系统优化升级为“深蓝”。1996 年，“深蓝”在费城对阵卡斯帕罗夫时只赢得了首局，卡斯帕罗夫宣称剩下的几局他赢得很轻松。次年，“深蓝”和卡斯帕罗夫纽约再战，意想不到的事情发生了。史上最出色、最令人敬畏的国际象棋大师加里·卡斯帕罗夫竟然被一台计算机打败了。

开局阶段：独立思考能力更重要

和其他事情一样，国际象棋比赛也包含着 3 个阶段：开局、中局和残局。但这项运动还有一点不同之处，那就是每个阶段所要考验的智力和情绪掌控技巧是不同的，这便使得国际象棋比赛如同精神界的铁人三项比赛一样，考验精神的速度、力量和耐力。

国际象棋比赛的开局阶段，棋盘中部是空的，兵、车、象整齐列队，靠边站成两排，等待主人发号施令。而命令的可能性几乎是无限的。白子有 20 种开局方式，黑子也有 20 种接招方式，第一回合完整攻守算下来就有 4 000 种可能的对阵方式。第二回合结束后共有 71 852 种可能的对阵方式；第三回合结束后共有 9 132 484 种对阵方式。一局完整的国际象棋比赛中可能出现的选手对阵方式的总数极其庞大，估量起来都成问题，但是有些数学家将这个数字定为 10 的 1050。这简直就是一个天文数字：正如迭戈·拉斯金·古德曼所写，“国际象棋比赛对阵方式的可能性比宇宙中的原子数还多。”

比赛开局时，所有棋子都在棋盘上，可能性是最不可限量的，此时计算机看似力量最为强大。正如 IBM 网站在（计算机对阵卡斯帕罗夫的）比赛前吹嘘的那样，“深蓝”每秒能计算出 2 亿个位置。该网站还恶毒地说：“顺便提一下，加里·卡斯帕罗夫的估算速度大概是每秒 3 个位置。”如此看来，卡斯帕罗夫哪里还有机会赢呢？

实际上，长久以来国际象棋计算机在开局阶段的表现都相当差劲。尽管此时的可能性最为无限，但目标也同样最不明确。当一棵树上有 10 的 1050 根枝条时，每秒计算出 3 个位移和每秒算出 2 亿个位移一样，都是徒劳无功的，除非是有导向地利用这种计算能力。

无论是电脑棋手还是人类棋手，都需要把一局比赛分解成若干个中期目标，例如，吃掉对方一个兵，或是将对方一军。在比赛中局，一旦棋子在攻守中受限，或是双方相互威胁，就会有很多策略目标浮现。那么，棋手需要作的就是想出策略来实现这些目标，并预测出哪一种方法对剩下的棋子最有利。然而，国际象棋比赛开局走法的目标要抽象得多。于是，计算机就要与这些抽象的开放式问题进行周旋，而人类却了解启发法，比如“占领棋盘中部”或是“让兵有序进攻”，并且能设计出许

多有创意的手段来运用这些方法。

另外，对棋手而言，开局的走法通常是常规走法，人类可以凭借数百年来的宝贵经验选择最佳走子方案。尽管从理论上讲，白棋有 20 种开局走法，但最佳的开局走法只有 4 种，竞争激烈的国际象棋比赛中约有 98%的比赛都以其中一种方法开局。

人类棋手面临的问题是，计算机程序通过研究统计数据，能将上述知识系统化。国际象棋数据库包含了将近上万场比赛，并且和其他数据库一样，能为预测所用。IBM 的程序员所研究的事物包括，过去出现的每一种开局方式的使用频率，使用这些方式的棋手比赛时的表现到底有多强势，每一套开局走法最终带来的胜局、败局和平局数各为多少。计算机分析这些数据时采用的启发法，即使没有远远超越人类的直觉和经验，也很有可能平分秋色。IBM 网站曾评价“深蓝”计算机庞大的数据库：“卡斯帕罗夫不是在和计算机比赛，而是在和从前那些象棋大师的灵魂过招。”

因此，在 1997 年对阵“深蓝”的 6 轮比赛中，卡斯帕罗夫第一轮的目标是要将“深蓝”引出数据库，让它无所适从。卡斯帕罗夫的开局走法相当常见，将他的马走到棋盘上 f3 格的位置。“深蓝”则派出棋子象威胁卡斯帕罗夫的棋子马——它必然会走这步棋，因为数据库显示，历史上此种情况下走象能使白棋的胜率从 56%降到 51%。

然而，这些数据库依靠的假说，是认为卡斯帕罗夫接下来的反应会和数据库中所有的棋手一样——把棋子马撤回去。结果，卡斯帕罗夫并没有理会棋子马被吃的危险，他认为“深蓝”只是在虚张声势，于是，他派出一个兵开路，这样他自己的棋子象就能控制棋盘中心。

卡斯帕罗夫的走法颇具战略意义，同时还实现了另一个目标。卡斯帕克斯已经走了 3 步，而“深蓝”只走了两步，此刻的对阵局面（见图 9–2）在“深蓝”程序背后上万场大师级比赛的数据库中仅仅出现过一次。

尽管采用的都是十分普遍的走法，但由于这棵大树上的枝条太多，走了 10~15 步棋之后，数据库能发挥的作用就不大了。若是一场“拉锯战”，很可能双方最终得到的棋局是国际象棋史上从未出现的。但是，卡斯帕罗夫只走了 3 步就跳出了数据库。本书已经提到，数据样本不足时，纯粹的统计学方法对预测的帮助不大。



图 **9–2** 第一局比赛中卡斯帕罗夫走完第 **3** 步棋后的棋局

“深蓝”需要的是独立思考能力。

中局阶段：宽度与深度的两难选择

在国际象棋比赛的中间阶段（简称为中局），计算机的优势可能就会逐渐凸显出来。此时，棋子可以自由移动到棋盘中部，每个回合可能的走法平均为 40 种，而不仅仅是开局时的 20 种。乍一看两者之间的差别不大，但随着可能性呈指数型递增，差别会迅速扩大。举例来说，假设你想提前计算出 3 个完整的回合（即，你和对手各走 3 步，总共 6 步）。开局时，需要计算的函数是 20 的 6 次方，即需要考虑到 6 400 万种可能性，这已是一个巨大的数字。而到中局时，则需要计算 40 的 6 次方，也就是有 41 亿种可能性。所有这些可能的局面，深蓝可以在 20 秒内计算出来，但是卡斯帕罗夫就算不吃不睡不洗澡，夜以继日地计算，也需要花上 43 年的时间才能算完。

像卡斯帕罗夫这样的伟大棋手从不会自欺欺人地认为自己能够计算出所有的可能性。这正是大师级棋手和业余棋手的差别所在。荷兰心理学家阿德里安·德·格鲁特曾经针对国际象棋选手做过一项非常有名的研

究，该研究发现，业余棋手下棋时总想找到完美的一招，寻求未果便会倍感沮丧，往往举棋不定，寸步难行。

国际象棋大师们则不然，他们并不刻意追求完美招数，而是着眼于一个还不错的棋步——当然，如果某个特定位置有绝妙的一步棋可走，他们也不会放过——他们会花费更多精力考虑如何走好这步棋，而不是绞尽脑汁地列出所有的可能性。美国大师级棋手鲁本·法恩曾写道：若指望国际象棋选手能够提前算出 20 或 30 个回合的步数，无异于“天方夜谭”。

这种情况并不是一句“完美是优秀的敌人”就能概括的。如果你想成为真正的国际象棋大师，则需超越这种简单的启发法。然而，若时间有限，且信息量超出了我们的处理能力范围，我们就无法做出完美的决策。但是，无论是在国际象棋比赛中还是其他涉及预测的活动中，承认这种不完美，也许反倒能让我们自由地做出最优选择。

这么说并不意味着像卡斯帕罗夫这样的象棋大师就不需要作任何计算了。事实上，为实现一些短期目标，如吃掉对方一个子，卡斯帕罗夫至少需要提前计算出 3~5 步可能的精确走法，从而形成策略。而每一种走法，他还需要想出对手的反应——比赛时可能会出现的所有招数——考虑对手的招数是否会破坏自己的策略。另外，还需要考虑到对方是否设置了陷阱；一个原本强大的阵形，若稍不留神未能及时保住棋子王，很可能几步棋之后就会被“将军”。

凭借记忆和经验，国际象棋选手懂得在哪一步上应当着力思考。有时，他们需要对这棵树上的很多枝条进行仔细研究，其实最终也就推出几步招数；其他时候，他们会关注某一根枝条，但会对某种走法进行深度推算。我们面对复杂问题时，总要在深度和广度之间做出权衡。例如，美国国防部和中央情报局就面临着这样的权衡，广撒网追踪更多的信号，以预测潜在的恐怖袭击；还是专注那些符合最邪恶分子特征的人，两者之间，他们必须作个选择。国际象棋选手中的精英比较擅长元认知，反思自己的思维方式，若无法实现平衡，他们就会进行自我修正。

国际象棋计算机可以在某种程度上兼顾宽度和深度。这些机器使用启发法来修剪它们的搜索树，将处理能力更多地放在更有发展前景的枝条上，而不是对每一种可能性都进行同等程度的计算。但其实，它们比人类计算速度快得多，所以不需要像人类那样做出太多妥协和让步，就可以对所有可能性稍作衡量，然后再对那些看似最重要的可能性作更详细的衡量。

但是，国际象棋计算机程序通常没有大局观，也不会作战略性思

考。它们非常擅长对战术进行估算以实现某个近期目标，但却不善于判断哪个目标在整个比赛方案中是最重要的。

卡斯帕罗夫试图引诱“深蓝”进行无战略意义的盲目追捕，以找到其启发法的盲点。国际象棋计算机程序通常倾向于短期目标，这些目标可被分解和量化，并且不必把整盘棋当作一个有机整体来评估。有关计算机的偏好，有个经典的例子——它乐意接受牺牲；当一个强手愿意以一个更有用的棋子来换一个用处不大的棋子时，它会欣然接受。

“当对手以牺牲更强大的棋子作交换时，接受这笔交易”，这个启发法通常是很奏效的。但当你对阵的是卡斯帕罗夫这样的对手时，这个策略未必会奏效，卡斯帕罗夫愿意作为交易中看似较弱的一方；卡斯帕罗夫明白战略得利比战术损失更重要。卡斯帕罗夫在第一局比赛用 30 步棋给“深蓝”设了一个局：他为保住棋子象牺牲了车这个棋子。看到“深蓝”中招，卡斯帕罗夫很高兴。当时的局面如图 9–3A，显示了计算机因缺乏战略性思考而造成的盲点。

从图 9–3A 中，可以看出卡斯帕罗夫和“深蓝”对简化棋局各有其道。计算机把复杂的问题分割成许多离散元素。比如，对“深蓝”来说，目前的棋局更像图 9–3B 所呈现的那样，每颗棋子代表不同的分值。若将各个棋子的分值相加，“深蓝”会以一个兵的优势战胜卡斯帕罗夫，这个棋局意味着在大多数情况下，“深蓝”会获胜或者打成平局。

相比较而言，人类更能够专注于最重要的元素，能从战略大局出发考虑问题，而战略整体功能有时会大于各部分之和。对卡斯帕罗夫来说，目前的棋局就像图 9–3C 显示的那样，确实是有利于自己的局面。卡斯帕罗夫看到他有 3 个兵在向“深蓝”的棋子王进逼，而且这时棋子王的防御较弱。“深蓝”的棋子王要么让出一条路来——在这种情况下，卡斯帕罗夫的兵可以长驱直入“深蓝”的王宫，升变为棋子后——要么就等着被“将军”。与此同时，虽然现在卡斯帕罗夫的棋子后和棋子象在棋盘的左下角，但是却能成对角线活动，几乎不受阻碍，给兵临城下的“深蓝”的棋子王施加压力。卡斯帕罗夫尚不确定“深蓝”的棋子王会被怎样吃掉，但是他知道目前自己已胜算在握。显然，“深蓝”也明白目前的处境，在 13 个回合后便投降了。



图 **9–3A** 第一局比赛卡斯帕罗夫第 **32** 步棋后的棋局

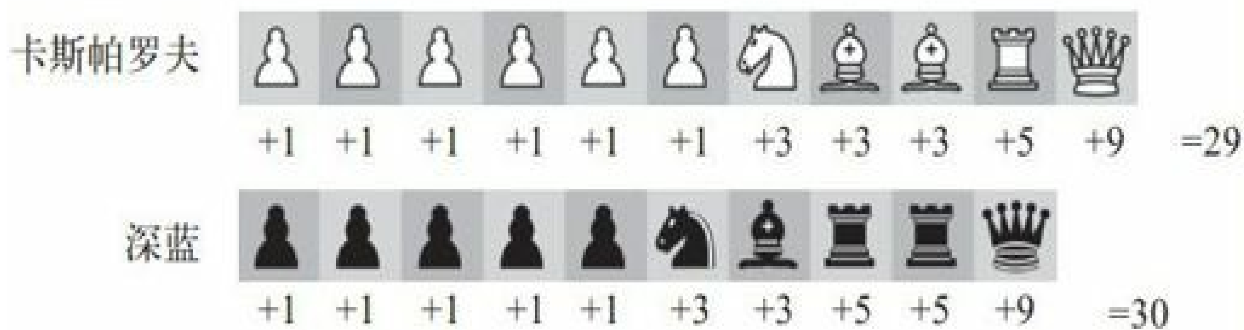


图 **9–3B** 棋局的离散评价

后来，卡斯帕罗夫说道：“这是计算机的典型弱点，我确信它对当时的局面很满意，对它而言这个局面的结局实在难以正确判断。”

“人类比计算机更擅长计算”，这是次日《纽约时报》的头条新闻。这一天，《纽约时报》发表了至少 4 篇关于这场比赛的报道。

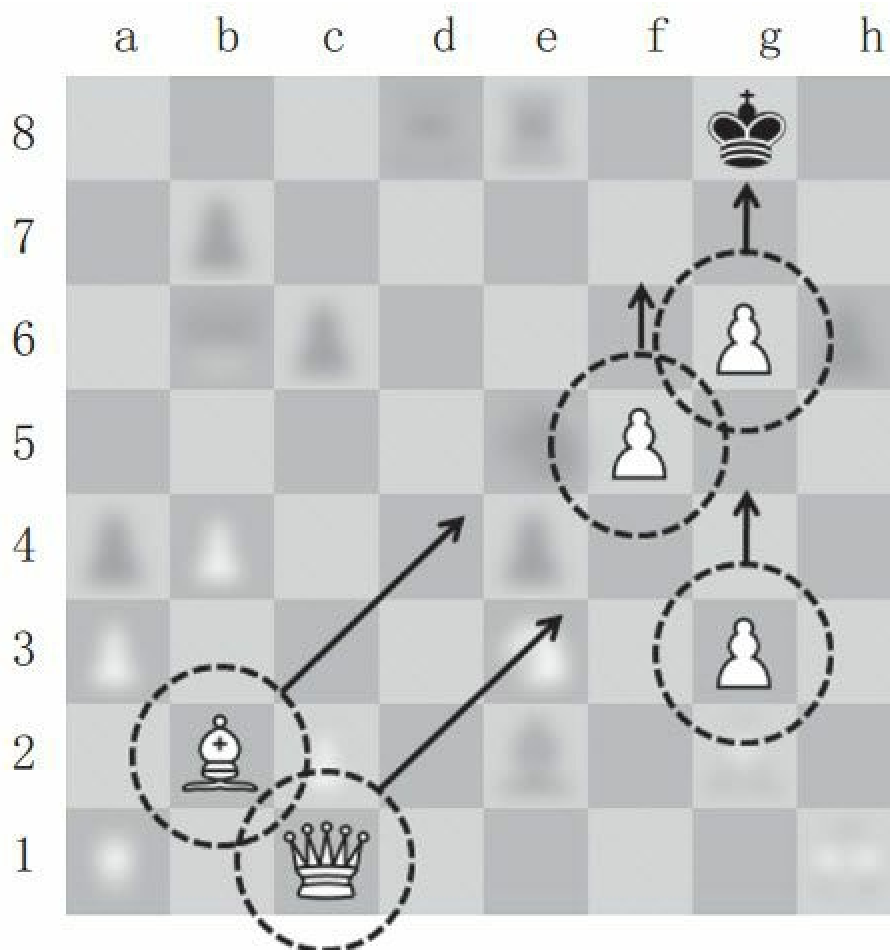


图 **9–3C** 棋局的整体评估

但是，比赛并没有最终结束。当时很多评论家并不太在意这场比赛，但它也许已经改写了国际象棋史。

残局阶段：计算机能力方面的较量

国际象棋比赛的最后阶段——残局，棋盘上的棋子越来越少，更容易计算出获胜的方式。然而，这个阶段仍需要缜密思考，因为所走的几十步棋必须精确无误，才能艰难地获得最终的胜利。举一个极端的例子，图 9–4 所示的棋局表明，不管黑棋怎么走，白棋都会赢。要走出这样的位置需要白棋准确无误地连续走 262 步。

人类棋手几乎不可能破解图 9–4 中的棋局。然而，人类有许多完成残局的实践，通常需要走 10 步、15 步、20 步，或者 25 步来完成。

残局对于计算机来说，是好坏参半的事。此时，几乎不需要短期战术目标，除非它能迅速结束残局，否则就容易因小失大。然而，计算机不仅有开局的数据库，也有残局的数据库。剩 6 个或更少棋子的棋局，理论上都能被破解。而 7 个棋子的残局则会复杂很多——有的解法复杂

到需要走 517 步——但是，计算机数据库里已经存储了怎样走会赢、会输或和局的步法。

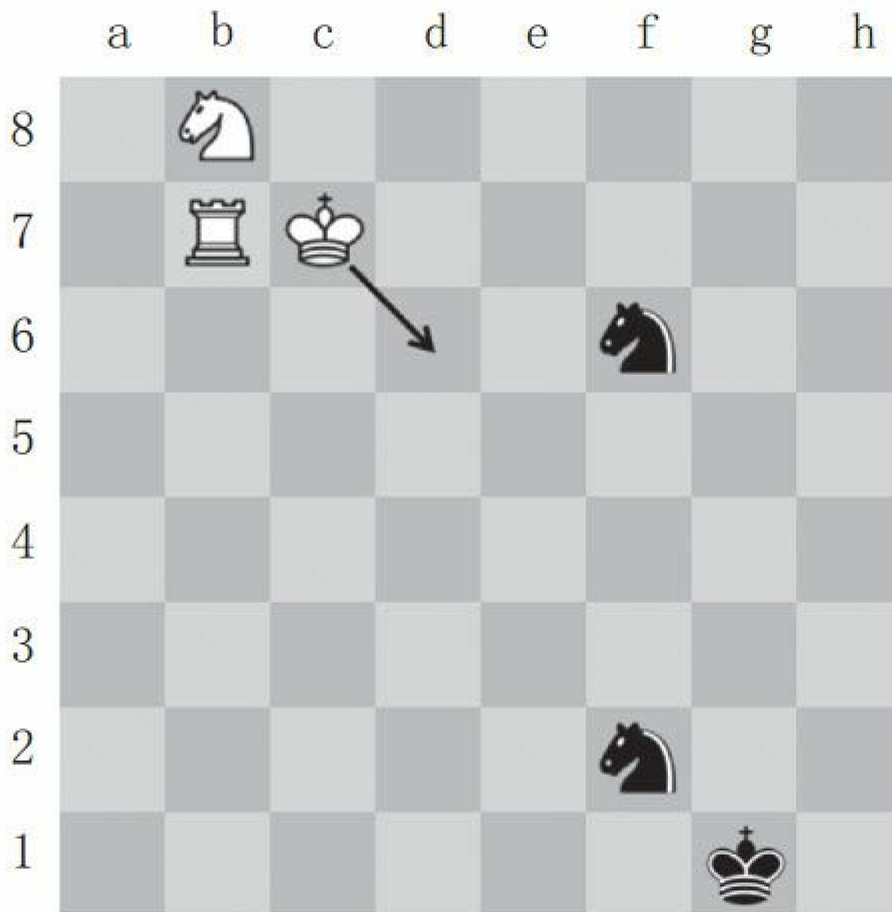


图 **9–3C** 棋局的整体评估

于是，这一阶段的比赛会出现某种类似黑洞的东西：当所有该和的局都和了，所有该赢的局都赢了，在“对策之树”（即所有策略的集合）的重力之外就不可避免地会出现一个点，这个点就类似于黑洞。国际象棋比赛中局的抽象目标被一些具体目标取代，比如，把自己后翼的兵派到这里你就赢了，引诱黑棋的棋子车移向其他地方就会和局。

在对阵卡斯帕罗夫的第一场比赛中，“深蓝”的系统显示当时的输局已定，但它还想继续与卡斯帕罗夫对弈，因为即使是卡斯帕罗夫这样的大师，每走 75 步也会犯一次严重的错误。而仅这样一步错棋就可能触动“深蓝”的传感器，进而让它找到取得平局的招数。“深蓝”计算机的处境不佳，但绝不是毫无希望。

“深蓝”非但没有退却，反而做出了一些非常奇怪的举动，至少在卡斯帕罗夫看来是这样的。在第 44 步棋的时候，“深蓝”把它的一个棋子车走到白方的第一行，而不是选择常规套路，将对方的军。计算机似乎

是毫无目的地乱走棋。在四面楚歌的时候，看起来“深蓝”大势已去，还放卡斯帕罗夫的兵进入己方第二行，随时威胁己方的棋子后。更为奇怪的是，此后只走了一步棋，“深蓝”就认输了。

卡斯帕罗夫好奇计算机到底是怎么想的，他看惯了“深蓝”犯战略性错误——比如，愿意用棋子象和棋子车作交换——在复杂的棋局中，“深蓝”根本不能深刻地认识到每步棋带来的影响。但是，在相对简单的棋局，计算机绝对不会犯战术错误。



图 **9–5** 深蓝使人困惑的一步棋

当晚回到广场大酒店回顾比赛的时候，卡斯帕罗夫问他的朋友、一位来自德国的象棋记者兼计算机专家弗雷德里克·弗里德尔：“计算机怎么会以这种方式自杀呢？”有些解释看上去似乎很合理，却没有一条能让卡斯帕罗夫觉得很满意。也许“深蓝”的确是“自杀”了，因为它发现无论怎么布阵走棋，这局都赢不了，索性认输，让卡斯帕罗夫摸不透自己的棋路。卡斯帕罗夫推测，难道这是一种煞费苦心的布局？也许程序员故意让“深蓝”输一局，从而使傲慢的卡斯帕罗夫过于自信进而导致失败？

卡斯帕罗夫的反应再自然不过了，他变得焦虑不安，开始研究起比

赛数据来。在弗里德尔和计算机国际象棋弗里茨的帮助下，卡斯帕罗夫发现，对计算机“深蓝”而言，传统的走法——黑方的棋子车走到第六列后将白方一军——根本不是好棋，这步棋最终会给卡斯帕罗夫一次将军的机会，尽管仍需 20 多步棋才能实现这一目标。

但这其中的深意是相当令人震惊的。按照弗里德尔的推断来看，唯一让计算机继续某种走法的原因就是它发现了另一种走法，这种走法会让卡斯帕罗夫走更多的棋步才能将“深蓝”一军，而原来的走法他只需要走 20 步就能“将军”。据弗里德尔回忆：

> 计算机“深蓝”其实已经算出了比赛的结局，它只是选择了损失最小的走法。卡斯帕罗夫说：“这台计算机已经算出来我再走至少 20 步就要将军了。”他为自己对多数计算保持正确判断而倍感欣慰。

人们曾经认为，让人类和计算机在象棋这种复杂比赛中提前算出 20 步棋是不可能的。卡斯帕罗夫称，他最引以为傲的时刻是 1999 年在荷兰参加的一场比赛，他提前算出了 15 步赢棋的步数。此前，人们普遍认为“深蓝”最多只能算出 6~8 步。而此刻，卡斯帕罗夫和弗里德尔不确定到底发生了什么事情，但是对那些漫不经心的观察家而言貌似很偶然、很莫名其妙的失误，对这两位来说则好像展现了伟大的智慧一般。

卡斯帕罗夫可能再也无法战胜“深蓝”了。

计算机战胜了人类

在第二局比赛中，“深蓝”更加咄咄逼人，丝毫不让卡斯帕罗夫有喘息的机会。第 35 个回合是关键环节。当时双方势均力敌：各有一个棋子后、一个棋子象、两个棋子车和 7 个棋子兵。但“深蓝”操纵着白子，略占优势：下一步是它先走，而且它的棋子后有足够大的回旋余地。目前的局面（图 9–6）对卡斯帕罗夫并没有太大威胁，但他的劣势是：接下来的几步棋将打开局面，决定“深蓝”是否有机会赢棋或者转向平局。

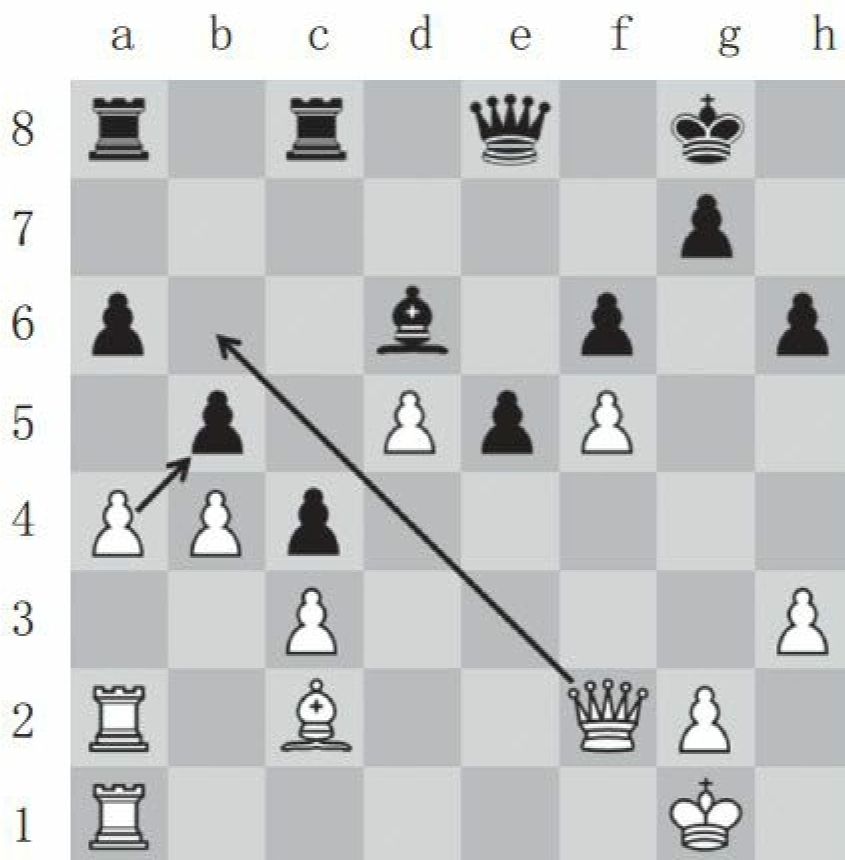


图 **9–6** 第二局比赛第 **36** 步棋后“深蓝”的选择

“深蓝”可以走的棋路很多，它可以让棋子后走到更具敌意的位置，这是比较有谋略的走法。或者可以以兵抵兵，打开左路棋局，这会使棋局更简洁开阔，是富有战略意义的走法。

现场比赛的评论员一致认为“深蓝”会采取第一种走法，推进棋子后。这是比较明显的走法，也更符合“深蓝”的特征：计算机更喜欢计算复杂的走法。但是，经过反常的长时间“思考”后，“深蓝”选择了以兵抵兵。

顷刻间，卡斯帕罗夫释然了，因为以兵换兵能缓解他当前的压力。但是，随着对棋局评估的深入，卡斯帕罗夫越来越不安，他把头埋进臂弯，并且咬手指关节部位——一位观众说看到卡斯帕罗夫在哭泣。为什么“深蓝”没有把棋子后推向前线？的确，“深蓝”的这步棋没有明显的劣势——卡斯帕罗夫甚至可以想象换作人类对手在这种情况下也会走这步棋，比如卡斯帕罗夫长期的对手阿纳托里·卡尔波夫就会这样做。但是，计算机这么做需要有较好的战术理由——卡斯帕罗夫找不出这个理由到底是什么。除非卡斯帕罗夫的推测是正确的——“深蓝”能够提前算出至少 20 步棋。

卡斯帕罗夫和“深蓝”的对弈已经打了 8 个回合。对于观赛的记者和专家来说，胜负已见分晓，打防守的卡斯帕罗夫一开始就注定没有赢的机会，但他仍有机会打成平局。出乎观众意料的是，卡斯帕罗夫走了 45 步之后就认输了。卡斯帕罗夫知道计算机不会算错，若能提前算出 20 步的棋，计算机是不会犯错的。卡斯帕罗夫知道，“深蓝”即将赢得比赛，所以何不养精蓄锐迎战接下来的 4 局比赛呢？

观众席上爆发出强烈的掌声：这局比赛比第一局精彩多了。如果观众没有像卡斯帕罗夫那样想到来自“深蓝”的“将军”是不可避免的，那他们对棋局的思考肯定没有卡斯帕罗夫想得深。观众对“深蓝”其实是真心佩服的：它能像人类选手一样下棋。女子国际象棋世界冠军苏珊·波尔加向《纽约时报》称赞“深蓝”：“棋下得太漂亮了！”而乔尔·本杰明是参与“深蓝”团队技术方面的国际象棋大师，他也赞美道：“计算机‘深蓝’走出了冠军的风格，就像卡尔波夫。这不是计算机游戏，而是一场真正的象棋比赛。”

当晚，卡斯帕罗夫匆忙地离开公平中心大厦，没有接受媒体的采访，但他却将同辈大师们的评价铭记于心。或许“深蓝”就是“实实在在”的人，只是没有呈现为人形而已。或许“深蓝”就像两个世纪前的亚马逊土耳其机器人，有一位国际象棋大师偷偷摸摸地在幕后操控它的齿轮。又或许曾经和卡斯帕罗夫打过平手的国际象棋大师本杰明不仅帮助提高“深蓝”的棋艺，而且实际上还介入了比赛。

国际象棋冠军们对模型有着敏锐的嗅觉，像偏执狂一样寻找各种模型。卡斯帕罗夫在第二天召开的新闻发布会上指责 IBM 公司具有欺骗行为。卡斯帕罗夫认为计算机游戏就如同“马拉多纳口中的上帝之手”。1986 年世界杯阿根廷对阵英格兰的那场比赛中，著名球员迭戈·马拉多纳手球得分。录像显示，马拉多纳不是用头而是违规用手将球送进球门。赛后，马拉多纳更是自称这个进球靠的“一半是马拉多纳的脑袋，一半是上帝之手”。同样的，卡斯帕罗夫认为是智囊团在帮助“深蓝”。

当然，卡斯帕罗夫关于“深蓝”运转方式的两种推测是自相矛盾的，正如埃德加·艾伦·波关于亚马逊土耳其机器人的构想一样。“深蓝”下棋下得太好，简直不像一台计算机；又或许它智力超凡，人类根本无法望其项背。

然而，第二局比赛的退出对于卡斯帕罗夫来说却是一个错误。第二天吃午餐时，卡斯帕罗夫最信赖的助手弗里德尔和尤里·多科扬对他说“深蓝”其实是险胜。弗里德尔和尤里·多科扬连夜在弗里茨程序上对该棋局进行分析，结果显示，按照某种走法，只要再走 7 个攻守回合就可以“将”“深蓝”的棋子王，最终就能打成平局。卡斯帕罗夫眼神空洞地

望着第五大道，说：“说完了吗？这台计算机棋艺高超，我知道自己根本就没有赢的机会。”

虽然现在的比赛结果是平局，但卡斯帕罗夫的信心发生了强烈的动摇。有生以来，卡斯帕罗夫还从未输过一场国际象棋比赛，可现在却面临困境。更糟的是，他犯下了国际象棋界最严重的错误，那就是放弃原本可以平局的比赛，这是一个令人尴尬且从未有过的错误。观赛的记者和象棋大师都不记得上一次世界冠军犯这种错误是什么情形了。

卡斯帕罗夫步步紧逼的棋风帮助他赢得过世界冠军，但他认为这样的风格赢不了“深蓝”，遂决定采用非传统的方式，谨慎地诱骗“深蓝”，就像黑客套取程序漏洞一样。虽然卡斯帕罗夫第三局比赛的开局很不寻常，将“深蓝”诱骗出数据库，但这样的开局带来的最好结果就是平局。卡斯帕罗夫在第四局和第五局比赛中发挥得更加出色，似乎力道都用到了点儿上，但是仍对付不了“深蓝”强大的残局数据库，这两局比赛的结果都是平局。此时双方各胜一局，平三局，还剩一局决胜局。

决胜局当天，卡斯帕罗夫在公平中心大厦现身时面容疲倦，孤独迷茫。弗里德尔后来回忆说，从未见卡斯帕罗夫如此消沉。卡斯帕罗夫这局执黑子，他选择以卡罗–卡恩防御方案开局。人们通常认为卡罗–卡恩防御方案胜算较低——历史上黑方以这种方案开局的胜率仅为 44.7%——尽管卡尔波夫这类熟知这一方案的棋手完全能够提高胜率。可是，卡斯帕罗夫并不熟悉这种开局方案，在以往的国际象棋锦标赛中他几乎没用过这种方案。走了几步棋后，卡斯帕罗夫变得很紧张，连常规的走法都要思考很长时间。在走第 7 步时，他犯了一个严重的错误，过早地折了一个棋子马。卡斯帕罗夫立即意识到自己的错误，重重地坐到椅子上，丝毫不掩饰内心的郁闷。仅仅走了 12 步后——此时开赛还不到 1 个小时——卡斯帕罗夫就弃赛了，随即离开了比赛现场。

“深蓝”赢了。只是它赢后没有砰然作响以示庆祝，只是发出逐渐微弱的机器运行声。卡斯帕罗夫是不是因为选择了自己并不熟悉的开局方案而倍感疲惫呢？还是如国际象棋大师帕特里克·沃尔夫所言，卡斯帕罗夫是故意输掉比赛，让“深蓝”赢得名不正言不顺？卡罗–卡恩防御方案是卡斯帕罗夫的对手卡尔波夫常用的开局手法，卡斯帕罗夫选择它开局有什么特殊意义呢？

但是，很快人们就对这些细微之处表现得没有那么关心了。人们只是关心是机器战胜了人类。就像计算机 HAL 9000 接管了飞船一样，人们无比兴奋。此时就像《爱将我们撕裂》这首歌第 13 秒的情形一样，合成音乐的声音盖过了吉他即兴演奏，摇滚乐不再那么流行。一个新的时代开始了。

所有这一切已成定局。卡斯帕罗夫已经成为人类脆弱情感的受害者，一个微不足道的软件漏洞更让这种情况雪上加霜。

国际象棋大师为何败给了“深蓝”计算机？

“深蓝”诞生于 IBM 的托马斯·沃森研究中心——从威彻斯特县山麓看，这个研究中心就像一座美丽的月牙形复古建筑。该中心的大厅中放置了早期“计算机”的复制品，如查尔斯·巴贝奇设计的众多机器。这座建筑外表看上去有些锈迹斑斑，但是很多科学家把家安在这些办公室中，其中包括数学家伯努瓦·曼德勃罗，他也是诺贝尔经济学奖和物理学奖的获得者。

2010 年春，我去托马斯·沃森研究中心拜访了默里·坎贝尔，这位加拿大人性情温和，像个大男孩。早在卡内基梅隆大学开始“深思”这个项目时，坎贝尔就是该项目的主要工程师之一（他现在是 IBM 统计建模部门的主管）。坎贝尔的办公室里有一张卡斯帕罗夫的大幅海报，海报中的卡斯帕罗夫怒视棋盘，标题是：

且看计算机棋手如何马失前蹄？

卡斯帕罗夫对弈“深蓝”

1997 年 5 月 3 日~1997 年 5 月 11 日

但最终失利的却是卡斯帕罗夫，而非计算机“深蓝”，尽管这个结果与坎贝尔及其团队的预想大相径庭，但它却实实在在地发生了。

“深蓝”的设计初衷就是要打败卡斯帕罗夫，这个程序就是为打败卡斯帕罗夫而专门设计的。该团队试图预测卡斯帕罗夫最有可能采用的开局方式，并研究如何反攻。（卡斯帕罗夫使用以往在国际象棋锦标赛中很少使用的开局来避开他们的圈套。）在 1996 年与卡斯帕罗夫的对弈中，“深蓝”表现一般。另外，该程序在与思路相似的棋手博弈时，也出现了一些问题。因此，“深蓝”团队将该程序的处理能力提高了一倍，对其启发法作了进一步完善。坎贝尔知道计算机“深蓝”还需要对数据库进行更深入（且更具选择性）的搜索，才能与卡斯帕罗夫的深度战略性思维相抗衡。同时，研究人员还改进了“深蓝”对复杂棋局的适应能力，使其优势得到充分发挥。

坎贝尔告诉我：“对计算机有利的棋局通常都是棋子很多的复杂棋局，这样就会有许多设定的走法可选。我们希望遇到战术比战略更重要的棋局。这样就可以花一些心思琢磨战术了。”

如此说来，“深蓝”便是空前绝后的更“人性化”的象棋计算机了。虽

然博弈论在象棋中的作用与它在信息不完整的游戏中（如纸牌）的作用程度不同，但在开局中可能却是一样的。一个稍微逊色的走子就会打乱对手的阵脚，使他数月时间的准备毁于一旦——如果对手知道如何应对的话，数月作无用功的人就是你了。但是，大多数计算机都想下出“完美的”象棋，而不是根据对手的状况调整比赛策略。“深蓝”却能像人类棋手一样比赛，并且会将局面扭转至坎贝尔眼中相对有利的局面。

1997 年时，卡斯帕罗夫的棋艺实在太高超了，当时的确应当好好想想该如何设计程序，让“深蓝”赢得比赛。

从理论上讲，对计算机可以下国际象棋这种情况进行编程并不难：如果让一个国际象棋程序无限期地进行搜索运算，那么全部 10 的 1050 种棋局都能得到解决方法。坎贝尔对我说：“有一个运算程序简单易懂，又能解决国际象棋中的问题，我也许半天就能写出这个程序，若给它足够的时间运转，便能够完成比赛。”可他又哀叹道：“可这样的计算，可能需要花费人类一生的时间来完成。”

若想让计算机棋手打败世界冠军，通常要作无数次索然无味的反复试验。设计者需要思考，是否应该让程序多注意残局阶段，而在中局阶段应少耗费时间，以达到表现上的平衡？是否有更好的办法能让电脑在开局阶段就能评估出舍马换象的价值？另外，有的棋路的确暗含“将军”的机会或能变成一个陷阱，但最终却不能帮助棋手赢得比赛，对这种棋路，程序需花多长时间才能识别出其无用性并且对其忽略？

通过调整这些参数并观察调整后的变化，坎贝尔对“深蓝”进行过多次试验。但有时“深蓝”还是会犯错，会出现一些奇怪的、出人意料的走法。在这种情况下，坎贝尔只得询问老程序员：新出现的走法究竟是程序的特征——是一个预示着其技巧正在提升的顿悟时刻，还是程序中的漏洞？

我一般会说这是程序中的漏洞。在更宽泛的预测环境中，当一个模型产生了一个意料之外或难以解释的结果时，通常是由漏洞造成的。把噪声误认作信号的情况太普遍，也太容易了。漏洞甚至能让最出色的预测者前功尽弃。

在本书上一章中为你介绍过那位身家百万美元的篮球赌客鲍勃·乌尔加利斯。有一年，乌尔加利斯决定要在棒球比赛中一试身手。他设计的模拟器建议把赌注压在费城费城人队，但这次乌尔加利斯赌输了。结果证明是因为模拟器 1 000 个行代码中有一个字母输入错误了：费城人队的主场球场市民银行球场是一个有助于防守和本垒打的小型球场，其编码为 P–H–I，但乌尔加利斯的助手却错误地把编码输成 P–H–1。仅一个错误的行代码就足以“淹没”程序中所有的信号，害得乌尔加利斯将赌金

压在了“噪声”上。这次失败让乌尔加利斯沮丧万分，从此彻底停用了这个棒球程序。

对坎贝尔而言，他们设计的“深蓝”程序在国际象棋方面早就比设计人员厉害百倍，这是坎贝尔面临的挑战，因为很有可能“深蓝”走了一步他们都没走过的棋，但他们却未必知道这是由于漏洞造成的。

“刚开始调试‘深蓝’时，每次它走了一步不常见的棋，我都会说，‘哦，一定出错了！’我们就会去探究，查看编码，最后弄明白问题所在。但随着时间的推移，我们越来越少这样做了。虽然当该程序继续出现奇怪棋局时，我们还是会去查看一下，但我们却发现它已经想出了一些人类很难搞懂的走法。”

在国际象棋史上，最著名的几步棋也许要数国际象棋天才波比·费希尔在 1956 年所谓的“世纪大赛”中走的那几步了。当时只有 13 岁的费希尔在对阵国际象棋大师唐纳德·波恩盖尔时，出现了两次戏剧性的弃子——一次是费希尔白白地牺牲一个棋子马，几步棋后，他又故意地让自己的棋子后失去防护，而让棋子象前进一步。这两步棋都是正确的，几步之后，费希尔的策略就使波恩盖尔损兵折将，自己的优势则越来越明显。然而，不管是当时还是现在，都很少有国际象棋大师仔细地考虑费希尔的招数。一直以来，棋手们都对“除非是置换对方的棋子后，或是可以立刻将对方一军，否则绝不能放弃自己的棋子后”这样的启示法深信不疑，之所以这样做，是因为这些方法屡试不爽，几乎从没出现过失误。


a b c d e f g h
8
7
6
5
4
3
2
1

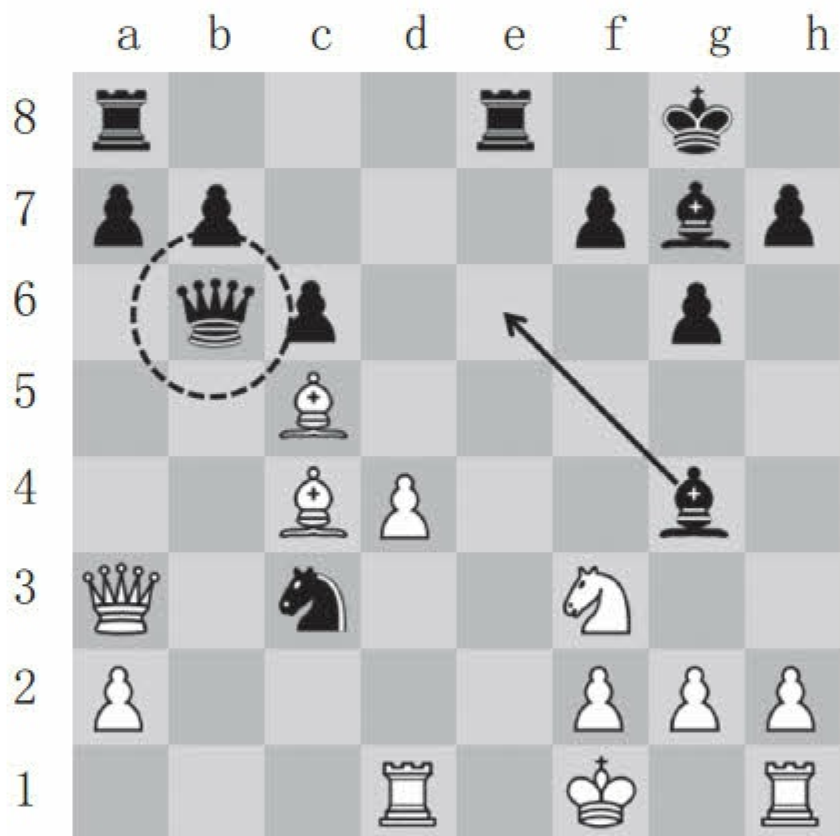

a b c d e f g h
8
7
6
5
4
3
2
1

图 **9–7** 波比 · 费希尔著名的牺牲招数（**1956** 年）

我把这个棋局输入我的笔记本电脑，通过费里茨程序运行，几秒钟之后，费里茨就识别出费希尔的招数。事实上，该程序认为这样的局面只有费希尔的招数方能奏效。在搜寻到所有可能的走法后，费里茨识别出这种情况下应该抛弃上面提到的那个启示法。

我们或许不能说计算机找到了这些走法是因为它具有“创造性”，相反的，它只是通过其强大的计算速度找到了这些走法。但是，计算机确实另有优势：下棋时，即使在走法上遇到了大难题，计算机也可以识别出特定棋局里的绝妙招数。但对于人类棋手而言，要做到这一点就需要突破传统，变得富有创造力和自信心。人们赞叹费希尔小小年纪却棋艺高超，可也许正是因为年龄小，他才能创造出这个招数：他充分地发挥自己的想象力处理每一步棋。人类的思维盲点通常是我们自己一手造成的，并且会与日俱增。计算机也有盲点，但它至少能够考虑到所有可能出现的招数，进而避免这些空想问题。

然而，“深蓝”的程序中还存在一些漏洞：数量不多，但肯定会有几个漏洞。采访接近尾声时，坎贝尔戏谑地提及 1997 年计算机“深蓝”和卡斯帕罗夫的对决，在第一场比赛快结束时发生了一个意外。

“‘深蓝’在比赛过程中出现了一个漏洞，这个漏洞很可能使卡斯帕罗夫对其能力做出错误判断，因为卡斯帕罗夫没想到这步棋是因为漏洞造成的。”

这个漏洞是在第一场比赛走到第 44 步棋时出现的，由于无法决定该走哪步棋，“深蓝”系统默认使出最后一招自动防故障装置措施，完全是随意走了一步棋。这个漏洞无足轻重，因为它是在比赛接近尾声时才出现的，而此时胜负已定。次日，坎贝尔及其团队修补了这个漏洞。坎贝尔告诉我：“1997 年年初，我们曾遇到同样的问题，当时认为漏洞已经修补好了，不幸的是，我们还是漏掉了一个漏洞。”

事实上，漏洞对计算机“深蓝”来说代表很多东西，唯独不代表灾难，因为漏洞似乎可以使计算机击败卡斯帕罗夫。卡斯帕罗夫迎战“深蓝”的比赛一直为人们所津津乐道，在第二场比赛中，卡斯帕罗夫出了一些状况，他犯了一个从未犯过的错误，输掉了原本可以打成平局的比赛。究竟是什么让国际象棋大师犯了这个错误？是由于“深蓝”第一局比赛中的第 44 步棋吗？当时“深蓝”似乎是极为随意地走了一个棋子车，这步棋让卡斯帕罗夫感到紧张，他觉得这有悖直觉的一步棋必是一步高招，而他绝没有想到这不过是一个程序漏洞。

因为太依赖 21 世纪的新兴科技，我们对机器在生活中的价值仍留有艾伦·波式的盲点。但计算机让卡斯帕罗夫出错，却是由于一个设计缺

陷。

计算机擅长做什么？

计算机的计算速度非常快。而且，它可以非常踏实忠心地计算——不知疲倦、不带情绪、不会中途改变分析方式。

但这并不意味着电脑做出的预测就一定很完美或者很准确。首字母缩略词“GIGO”（意为无用输入，无用输出）很好地概括了这一个问题。若给计算机输入错误的数据，或提供一套愚蠢的指令供其分析，它不可能“变废为宝”。此外，计算机也不善于完成需要创造性和想象力的任务，比如为这个世界的运转方式设计策略或提出理论。

因此，在气象预报和国际象棋等系统中，计算机对预测人员的作用最大，因为这些系统遵循的法则相对简单且容易理解，但是，为了做出准确的预测，还是必须反复计算支配系统的方程式。而在经济预测和地震预测方面，计算机的用处不大，因为我们对造成这些问题的根源所知甚少，而且这些领域的数据“噪声”又比较大。但 20 世纪七八十年代，这些领域对计算机主导的预测寄予厚望，当时的学者和科学家接触计算机的机会逐渐增多，但是此后并没有取得多大进步。

很多领域的情况介于这两种极端情况之间。数据通常还不错，但是仍不够好，对产生这些数据的系统和过程，我们都有所了解，但了解得还不够充分。在此类案例中，要想改进预测，也许能够借用“深蓝”的程序员们处理数据的方法：试错法。这也是许多公司经营战略的核心所在，这些公司和当今这个大数据时代可谓休戚相关。

用试错法提高计算机的预测能力

2009 年年末，我到加利福尼亚州芒廷维尤的谷歌总部进行采访，在那里，你很难分清谷歌的员工什么时候是在严肃工作，什么时候是在嬉笑玩闹。谷歌公司以“红、黄、蓝”三原色为装饰主色调，配备排球场，还有可想象到的各种造型的两轮摩托车，这样的公司文化能激发员工的创造力。每一个谷歌人，甚至包括工程师或经济学家，都可以不落俗套、异想天开。

“这类实验这里随时都有，”谷歌首席经济学家哈尔·瓦里安在和我会面时这么说道，“你更应该把这里看作一个有机体，一个有生命的组织。我曾经说过，某个物体被赋予生命后就应该得到我们的关注，就像《终结者》中的天网一样。但是，我们已经和加利福尼亚州政府达成协议。”那时，阿诺·施瓦辛格正担任州长。“他们会来帮助我们。”

谷歌对其研究项目和其他产品进行了广泛测试。瓦里安说：“2008 年，我们对研究项目进行了 6 000 次实验，广告货币化的实验大约也有

6 000 次。所以，谷歌一年大约要进行 10 000 次实验。”

其中有些实验是有形的——偶尔还能延伸出一条新的产品链。但大多数实验都较难被觉察：将公司的标志移动几个像素的位置，或是对广告牌上背景颜色的序列稍作改变，然后观察调整后的点击量的变化或是货币化的效果。其中许多实验会应用到约 0.5%的谷歌用户中，是否应用取决于这个构想的发展前景。

当你利用谷歌搜索词条时，你可能没有意识到自己已经参与到谷歌的实验中。但从谷歌的角度看，事情则稍有不同。谷歌返回给你的搜索结果和首页中显示的排序，代表了谷歌对你认为最有用的搜索结果所作的预测。

对有用性的衡量和预测貌似非常主观，那谷歌又是如何做到的呢？如果你搜索“最好的新墨西哥餐厅”这一词条，这是否说明你正计划到阿尔伯克基旅行呢？还是说明你在寻找最近刚开业的墨西哥餐馆？又或许你想要知道哪家墨西哥餐馆有怒希沃拉丁菜？也许你应该组织好搜索的问题，可你并没有这样做。于是谷歌会召集 1 000 个陪选团，他们都提出了相同的问题，谷歌会呈现给他们丰富多变的网页，让他们从零到 10 为每一个网页的有用性评级。然后，谷歌展示给你的网页就是按级别高地排列的结果。

当然，谷歌不可能对每个搜索问题都进行这样的操作，因为每天公司会收到上亿条搜索。但是，瓦里安告诉我，当出现有代表性的搜索问题时，他们仍会使用人工评测。然后他们会对比哪个统计测量的结果（尤其是从相关性和有效性上来看）和人工判断的结果相关。谷歌针对网站的统计测量最著名的就是 PageRank 信号了，这个测量案例根据网站的外部链接和内部链接的数量来衡量网站的价值。但是，PageRank 信号不过是谷歌 200 个信号中的一个，谷歌利用这些信号一步步向人工评估结果靠拢。

当然，这个任务并不轻松——要将 200 个信号应用到几乎存在无限可能的搜索问题中。这就解释了为什么谷歌如此强调实验和测试。你所知道的谷歌搜索这一产品质量已经很高了，但明天公司可能又会对其进行调整。

谷歌公司之所以成功，是因为它将严密谨慎的测试和随心发挥的创新文化有机地结合起来。公司鼓励员工超越电脑，去做电脑做不了的事情：要想点子，各种各样的好点子。随后，谷歌会使用大量的数据测试这些点子。其中大多数点子很快会遭到淘汰，但最棒的点子会保留下来。

计算机程序就是用这种筛选的方式下象棋，它能够探寻几乎所有可

能的选择，并进行一定程度的研究，但它更加关注那些更有攻击潜力的招数。这一点十分符合贝叶斯定理：谷歌永远都处于运行的开端，不断修正其搜索运算程序，永远不会因为觉得这些程序已经完成而停止。

谷歌基本上能够立刻获得全球百万名用户的反馈，但多数时候，我们无法像谷歌这样迅速获得反馈来检测一个构想。“深蓝”的工程师们可以方便使用超级计算机，但是我们却没有这个条件，所以进步速度就要慢得多。

然而，我们要关注自己的预测在现实生活中的表现，而不是只满足于它在数据模型中的表现，简言之，就是立下自我检查的承诺，这也许是加快我们学习预测过程的最佳途径。

克服人类的技术盲点

在许多方面，我们自身已成为最大的技术障碍。人类缓慢稳固的进化进程已经远远落后于科技进步的步伐：人类进化以千年来算，然而技术的处理能力每隔一年几乎就翻一番。

我们那些生活在山洞中的祖先可能已经发现了一个问题，即拥有强大甚至过盛的模式识别技能是极为有利的，这样就能够在一瞬间识别出远处树叶的瑟瑟声是风引起的还是一头入侵的灰熊制造的。如今，在这样一个快节奏的社会里，各种数字和统计数据泛滥成灾，原本的习惯和倾向反而让我们陷入麻烦：当我们看到一系列的随机数字时，竟能看出其中根本就不存在的模式。（广告人和政客们常常利用这种现代骗术来欺骗我们。）

国际象棋则能引导我们走向美好结局。卡斯帕罗夫和“深蓝”的程序员将彼此视为敌手，但是他们都教会我们一点——预测时，计算机的处理速度和人类思维的精巧应该相互补充。

事实上，目前世界上最精彩的象棋比赛既不是由某位人类棋手完成的，也不是某台机器完成的。2005 年，ChessBase.com 网站举办了一场“自由式”国际象棋锦标赛：选手可以随意地使用自己喜欢的计算机程序补充自己的见解，并通过互联网征求建议。尽管一些大师也参与了比赛，但是比赛的赢家既不是最厉害的人类棋手，也不是那些使用最先进电脑软件的选手，而是由来自新罕布什尔州的两名二十几岁的象棋业余爱好者——史蒂文·克兰普顿和扎克里·“柴克斯”·史蒂芬——共同获得，他们将 3 套计算机程序结合起来算出结果，共同决定棋路。克兰普顿和史蒂芬之所以获胜，是因为他们既不敬畏技术，也没有被它吓倒。他们了解每个程序的优缺点，他们更像一个教练，而不是选手。

然而，看到“电脑认为洋基队将赢得世界大赛”这样的句子时，你应

该保持警惕。但如果它是以下这句话的速记版本——“计算机程序输出的结果是洋基队将赢得世界大赛”——那它可能不会带来什么恶性后果。当今世界信息横流，能够拥有那些计算速度远超过我们的机器，肯定是大有裨益的。

但是，如果你感觉预测者更倾向于这种看法——认为计算机是一个有知觉的个体，或是一个有自己思想的模型——这也许就表明这位预测者根本就没有进行什么思考。无论预测人员带有什么偏见和盲点，这些偏见和盲点肯定会被复制到他的计算机程序中。

我们必须以客观的态度看待科技——它就是一个能改善人类生存环境的工具。我们既不应该对技术的祭坛顶礼膜拜，也不该被它吓唬住。至今还没有人设计出一台能与人类相媲美的计算机，或许永远也没有人能做到这一点。但计算机本身就是人类进步和智慧的反应：如果某个技巧是由人设计的，那它就算不得真正的“人工”智能。

第十章

扑克牌游戏：如何从 **1 326** 种组合中猜出对手的底牌？

2003 年，作为一种泡沫经济，“扑克牌游戏热潮”开始兴起，涌现出大量扑克牌游戏新手，即使只懂一点儿扑克牌技巧，也能大获成功。这一现象有两个直接成因，其一就是，在 2003 年拉斯韦加斯举行的世界扑克牌锦标赛上，一位 27 岁的业余选手拔得头筹。这位选手是一位来自纳什维尔的会计师，有一个非常吉利的名字，克里斯·“福星”（Chris Moneymaker）。“福星”简直就是广大普通扑克牌玩家的典型代表，他是一个微胖的办公室职员，初玩时只投下 39 美元的赌注，一路凭借胆子大、手气壮，最终竟然将 250 万美元收入囊中。

全球最大的体育电视网——娱乐与体育节目电视网（ESPN）把“福星”的成名故事制成了 6 个系列短片，几乎每天晚上都会重播，直到 NBA 填补这一时段的栏目空缺为止。在这之前，人们想到扑克牌游戏，认为它只有粗鄙、陈旧和令人恐惧的一面，可以说，这部系列短片极好地宣传了扑克牌“运动”。突然间，每个身高 1.77 米的秃顶会计都从“福星”的身上看到了自己的影子，这些人都早已放弃了成为下一个迈克尔·乔丹或德里克·杰特的梦想，他们与“福星”的身材差不多，从事的工作也类似，而“福星”能够在几周内从业余扑克牌选手变身为世界扑克牌锦标赛的最大赢家，自己又有什么不可以的。

ESPN 为观众呈现的是一个经过精心剪辑编排过的比赛，与扑克牌桌上的真实情况极为相似。一方面，总计 800 多位参赛选手参与的 40 多个小时的比赛，必须压缩成 6 个小时的节目播出，所以只播放了比赛实况的一些小片段。另一方面，借助安装在每张牌桌上的针孔摄像机，“福星”和其他每位选手的牌都能为观众所知，这让观众觉得自己好像拥有“千里眼”。如果知道对手拿的是什么牌，那么扑克牌游戏玩起来就相当简单了。

在节目中，“福星”扮演的是战无不胜的主角。冷静分析后，人们就会发现尽管“福星”打得并不好，但仍然被解说员称赞——虚张声势被视为勇敢之举，过早弃牌则被赞为有先见之明。“福星”并不是水平较高的赌徒，会终生与牌相伴，他只是一个精明的扑克牌游戏专家，几乎一夜之间就成了世界级的扑克牌游戏玩家。

这让许多人相信，扑克牌这项“运动”易学、易获利，并且充满了让

人难以置信的激动时刻，但事实上并非如此。这并不能阻挡人们冲向拉斯韦加斯的脚步，或许下一个克里斯·“福星”就要诞生了呢？参加扑克牌世界锦标赛的人数暴增，仅 3 年时间，参赛人数就从福星参赛那年的 839 人增加至 8 773 人。

我也是其中之一。我一度也幻想成为扑克牌玩家，不过后来幻想破灭了，因为我认识到扑克牌玩家是在信号与噪声混合的泥沼中摸爬滚打的。从扑克牌游戏中，我认识到概率在我们生活中的作用，当我们设法理解世界、预测世界时，概率会带来哪些错觉。

是心理游戏，也是数学游戏

扑克牌游戏热潮兴起的另一个催化剂是互联网。互联网扑克牌游戏在 1998 年已有了雏形，随着扑克牌聚会和扑克牌之星这类扑克牌游戏公司不断地开拓网络赌博市场，使其日渐合法化，到 2003 年时，互联网扑克牌游戏已成为主流游戏。来自世界各地的玩家克服了对虚拟桥牌室的安全性和合法性的顾虑，蜂拥而至。因为这种虚拟桥牌室为玩家提供了各种便利：各种类型的棋牌游戏全天候供应，每局的赌金从几美分到几百美元不等；能够体验飞一般的网速（电脑比人的洗牌速度要快得多）；在家上网玩牌比在烟雾缭绕、破损老旧的赌场要舒服自在得多。

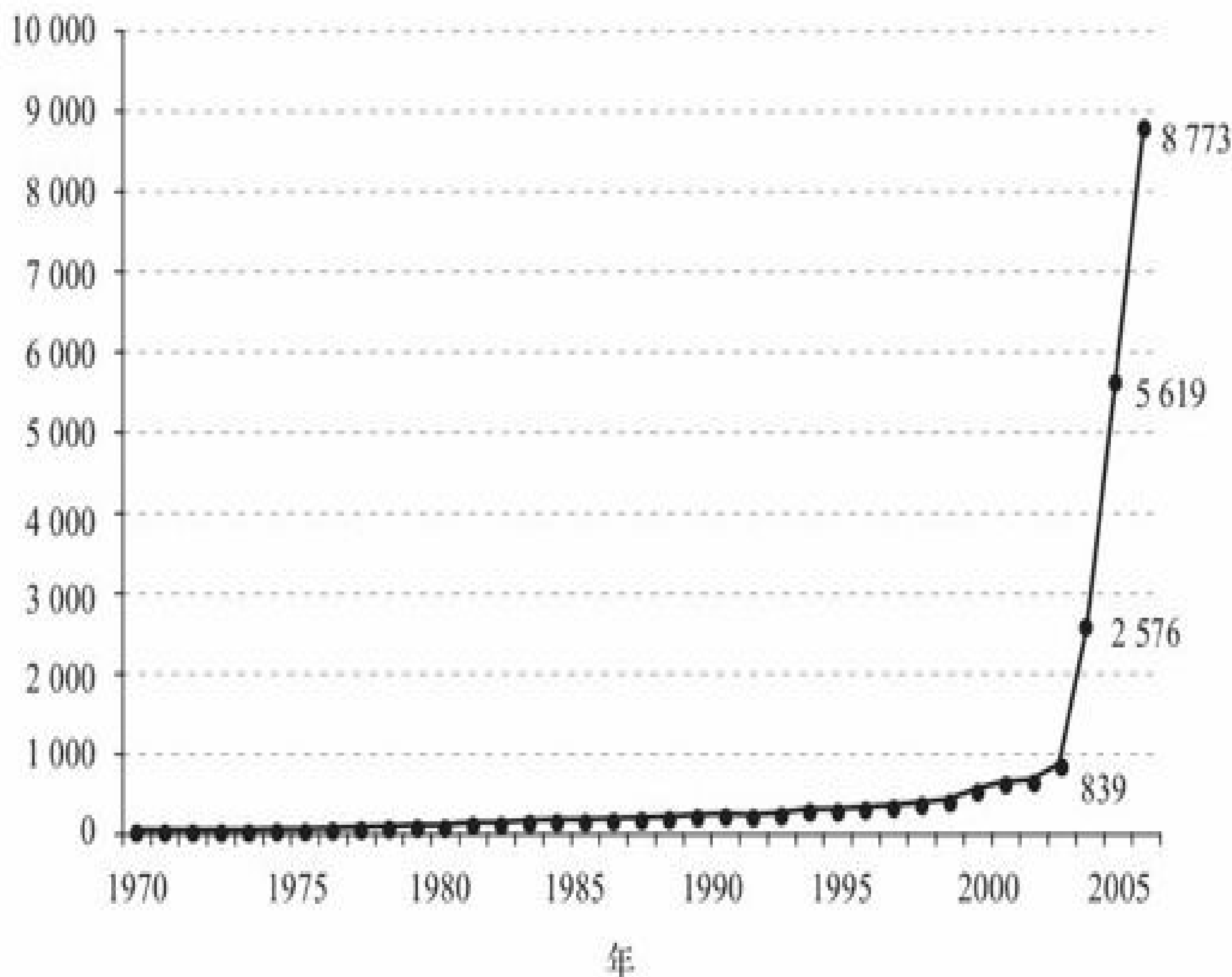


图 **10–1** 世界扑克牌锦标赛等主要赛事的参加人数（**1970~2006** 年）

前文提到，我曾经在毕马威会计师事务所担任经济顾问，工作与“福星”并无二致。一位关系很好的同事提出与我合伙设计一款常规游戏，因为赌博使我们的神经过于紧张了。曾经有那么几次，清晨 4 点钟我仍在密歇根欢喜山的“翱翔雄鹰”印度赌场逗留，所以还算有一点儿扑克牌游戏经验。但现在早已手生，急需实践，于是我也“转战”网上扑克牌游戏。一个名为“太平洋扑克”的怪异网站为玩家提供了一个难以拒绝的优惠：仅需 25 美元现金作为筹码，几乎没有任何附加条件。

我很快就输掉了最初的 25 美元，但与“翱翔雄鹰”赌场里那种既有刑满释放人员又有七旬老者的鱼龙混杂的场景相比，“太平洋扑克”网站上的棋牌游戏似乎还算简单。于是，我又为自己的游戏账号充值了 100 美元。几乎所有的职业扑克牌玩家都是以连胜局开启赌博生涯的——一开始就输的人通常会“聪明”地知难而退——我也一样。刚开始的时候，我每晚会赢 50 美元或 100 美元，甚至有时会赢 500 美元或 1 000 美元。

大概过了 3 个月，我的赌资冲破了 5 000 美元的大关，于是我整晚打牌，破晓时分就乘坐出租车去上班，工作时间感到浑浑噩噩。6 个月后，我赢了 15 000 美元，于是我辞职了，离开了令人兴奋的国际税务咨询行业的世界，此后我的时间一半用于打牌，一半用于撰写《棒球规程》。那段时光真是无拘无束，我甚至觉得自己好像已经破解了这个扑克牌游戏网站的系统。

一开始我还不知道自己是不是真正擅长打牌，但比赛的门槛很低，而我又有统计学背景的优势。扑克牌有时被视为一种具有高度心理学性质的游戏，是一场意志之战，因为双方都在通过凝视对方的心灵寻找暴露对方牌面的“话语”，试图读出对方的心思。这样说也不为过，尤其是在投注上限较高时更是如此，但也不像你想象的那样。（扑克牌中的心理因素大多以自律的形式出现。）其实，扑克牌更像一种数学游戏，与所有预测类型相同，这种游戏靠的是对不确定因素做出的概率判断。

出色的扑克牌玩家如何读出对手的牌？

优秀的扑克牌玩家之所以出众，并不是因为他们拥有超强的感知力，能够预知下一张牌是什么。只有极度迷信或患有妄想症的玩家才会认为洗牌定是有序可循，也只有一些牌技不佳的玩家才无法进行基本概率的计算：当你已有两张花色相同的牌时，同种花色的 5 张牌凑齐的概率约为 1/3，或者 80%的情况下会出现一对 A，而不是一对 K。然而，扑克牌的精髓在于玩家的“读牌”能力，这是一种分析能力，判断出对手手中的牌和他（她）的出牌套路。

这一技能极具挑战性，在人们最常玩的德州扑克中更是如此。在德州扑克中，玩家的底牌都是牌面朝下的，在所有玩家投下赌注之前，对手的手牌（即手中持有的牌）是无从得知的，因为每个玩家手上的牌都有 1 326 种排列组合，可能是一对 A 这样的大牌，也可能是 2~7 中任意两张小牌，但是，若玩家舍不得下赌资，那他的牌再好恐怕也玩不转这个牌局。

然而，玩家却可以利用读牌术预测对手手牌的范围，他们经常说“给对手上一手牌”，有时在玩牌的过程中，他们仿佛对对手的两张底牌了如指掌。但是，最出色的玩家通常喜欢做出许多假设，并用这些假设揣度对手的表现。一个合理的预测应当是概率性的，每结束一局，预测就应该更精准一些，但通常来说，只有当所有的牌都亮出时才有可能预测出对手手中的牌是 1 326 种可能性中的哪一种。若对手牌技高超、出牌谨慎，那就更难预测了。

在德州扑克中，信息的确难求，这使得牌手们在发牌之前就开始估

计对手的手牌范围。在线游戏中，这个过程通常是由数据分析完成的：你得到的数据显示出每个对手在前几轮比赛中的表现，他们在打牌的过程中是松懈还是紧张、被动还是主动。若是在真实的赌场中，则是通过牌手们彼此过招的历史记录来判断，如果没有交手过，那只能去查看他们的个人资料了。比如，人们发现来自瑞典、黎巴嫩和中国的牌手就比来自法国、英国或印度的牌手更善于主动进攻；而年轻的牌手就比年长的牌手更加松懈；男性牌手就比女性牌手更爱唬牌。这种思维定势也不总是对的，我在拉斯韦加斯百乐宫玩德州扑克时，最好的玩家以女性居多，她们之所以打得好，恰恰是因为她们比对手更喜欢主动出击。人们总认为女性的玩法更加保守，虽然这种观点存在 45%的错误概率，但是在 55%的情况下这种说法都是正确的，这一事实还是让他们费了一番脑筋的。

游戏一旦开始，粗糙的假设就会被更为可信的信息所取代：当天这位牌手在前几轮打得如何？这一轮又打得如何？每位牌手在每一轮下注、看牌、叫牌之后，都要更新其概率评估，这样的过程从根本上说，就是贝叶斯定理的套路。如果你曾怀疑过贝叶斯定理的实际应用，那可能是因为你从未玩过扑克牌游戏。

德州扑克速成法则

我们很容易从网上或书上查到德州扑克的游戏规则，但我还是要为初学者提供几条最基本的规则。首先，我要向你介绍一下下文中会出现的专业术语。德州扑克的游戏规则和其他牌类游戏一样，都很简单，但正如需要高超的战术和战略的国际象棋，规则简单并不代表游戏简单。

游戏开始时，每个玩家分别得到两张牌面朝下的个人牌（口袋牌，也称底牌），在这一阶段就要进行第一轮下注了，桌上有 5 张牌面朝上的“公共”牌，所有玩家都可以看到。每个玩家都在努力构想如何在底牌和公共牌之间搭配出最佳的 5 张牌。公共牌是依次发出的，每发一次公共牌，玩家就要下一次注。第一次同时发 3 张公共牌，称为翻牌（这是扑克玩家所使用的许多丰富多彩的词汇之一）。接下来发第四张公共牌，称为转牌。最后揭晓的是第五张牌，称为河牌，这时候就要进行最后一轮下注了。最后会有两种结果，大多数情况下，除赢家之外的所有玩家此时都会弃牌。如果不是这样，所有玩家开始亮牌比大小，成牌最大的玩家赢取底池。

起手牌范围如下：

同花顺（K♥Q♥J♥T♥9♥）

4 张/4 同点——4 张相同的牌(7♣ 7♥ 7♠ 7♦ 2♥)

满堂红——3 条加一对(Q♣ Q♥ Q♠5♦ 5♣)

同花(A♠ J♠ 9♠4♠2♠)

顺子(8♣7♦ 6♠ 5♦ 4♥)

3 条——3 张相同的牌(9♦ 9♠ 9♠ A♣2♥)

两对(A♥ A♠ 3♣ 3♦ 7♠)

一对(K♠ K♥ 9♣ 8♣6?)

散牌(A♣Q♠8♦ 5♦ 3♠)

如果出现平手牌，那么牌序较高的玩家获胜：比如，最大到 A 的顺子大于最大到 9 的顺子。如果是对子相同的情况，第三高的牌（起脚牌），即不成对的最大一张牌将会决定胜负，比如手牌(8♥8♠ K♣7♥ 5♥)大于手牌(8♥8♠ Q♦7♥ 6♠)，因为 K 大于 Q。

是弃牌还是跟注，如何选择？

假设你正在百乐宫里酣战，沉迷在一场 5/10 美元无注限的德州扑克牌当中[1]。前面几位玩家已经弃牌了，而你却发现自己的牌还不错，有一对 8（8♥8♣），于是你将赌注提高到 25 美元，而且只有一位玩家跟注，是一位六旬老者，我们称他为“律师”。

“律师”为人友善，在发牌的过程中很喜欢聊天，但在发牌结束后就相当安静了。我们了解到他在东海岸的一家知识产权法律事务所工作，而且干得不错。你可以想象“律师”穿着一件保罗衫，会定期发短信通知他的朋友查看高尔夫公开赛的开球时间。他喝了一杯啤酒，这时鸡尾酒服务员走过来将啤酒杯换成了一杯咖啡。“律师”不太可能被这些金额不大的赌资吓倒，他的表现也不像受到惊吓的样子。

当“律师”第一次坐到我们这一桌时，我们还在揣测他会如何出牌：第一，他可能会小小卖弄一番，为了把我们唬住而虚张声势；第二，他可能只是墨守成规，按部就班。我们随后的观察证实了第二种猜测可能更符合他的情况：他似乎是一个中等水平的玩家，一直在规避犯错，但也没什么技巧。他虽不是桌上牌技最差的玩家，但恐怕也不是能坚持到

最后的赢家。不过，我们还没有长期交手的经历，对他的秉性并不能完全确定。

那么，我们对“律师”的牌面又有多少了解呢？唯一能确定的，就是他的牌里肯定没有 8♥或8♣，因为这两张牌在我手里。不幸的是，起手牌的组合数只不过是从 1 326 种下降到了 1 225 种，而这 1 225 种组合中的每一种出现的概率都是相同的。

然而，“律师”的做法还向我们传递了另外一些信息：他决定跟注。大多数玩家，包括“律师”在内，在亮牌之前如果不能加注，都会选择弃牌。“律师”的这种做法说明他的牌还不错。尽管存在这种可能性，但他的表现也说明这幅手牌不可能特别好，里面不会有一对 A 这么好的牌，因为他既没有弃牌，也没有加注[2]。

我们可以运用贝叶斯定理来推测“律师”的手牌。从以往的经验来看，这样的玩家手中的底牌很可能是对子，如 9♦ 9♥，也可能有一张A，或者有两张牌的花色还相同，如 A♥ 5♥，这就意味着这副牌更容易凑成同花。同样的，他的底牌可能是 6♠ 5♠这样两张花色相同的连续牌，既可凑成同花，也可凑成顺子，这就是人们所说的“同花连牌”，又或者他的底牌包含的是 K♣ J♦这样的“人头”牌。

如果时间充裕，与能够迅速列出各种可能性的计算机一样，我们也能列出“律师”手牌的各种可能性，根据他目前的叫注，我们可以给 1 326 种可能赋予零至 100%不等的概率，如图 10–2 所示。

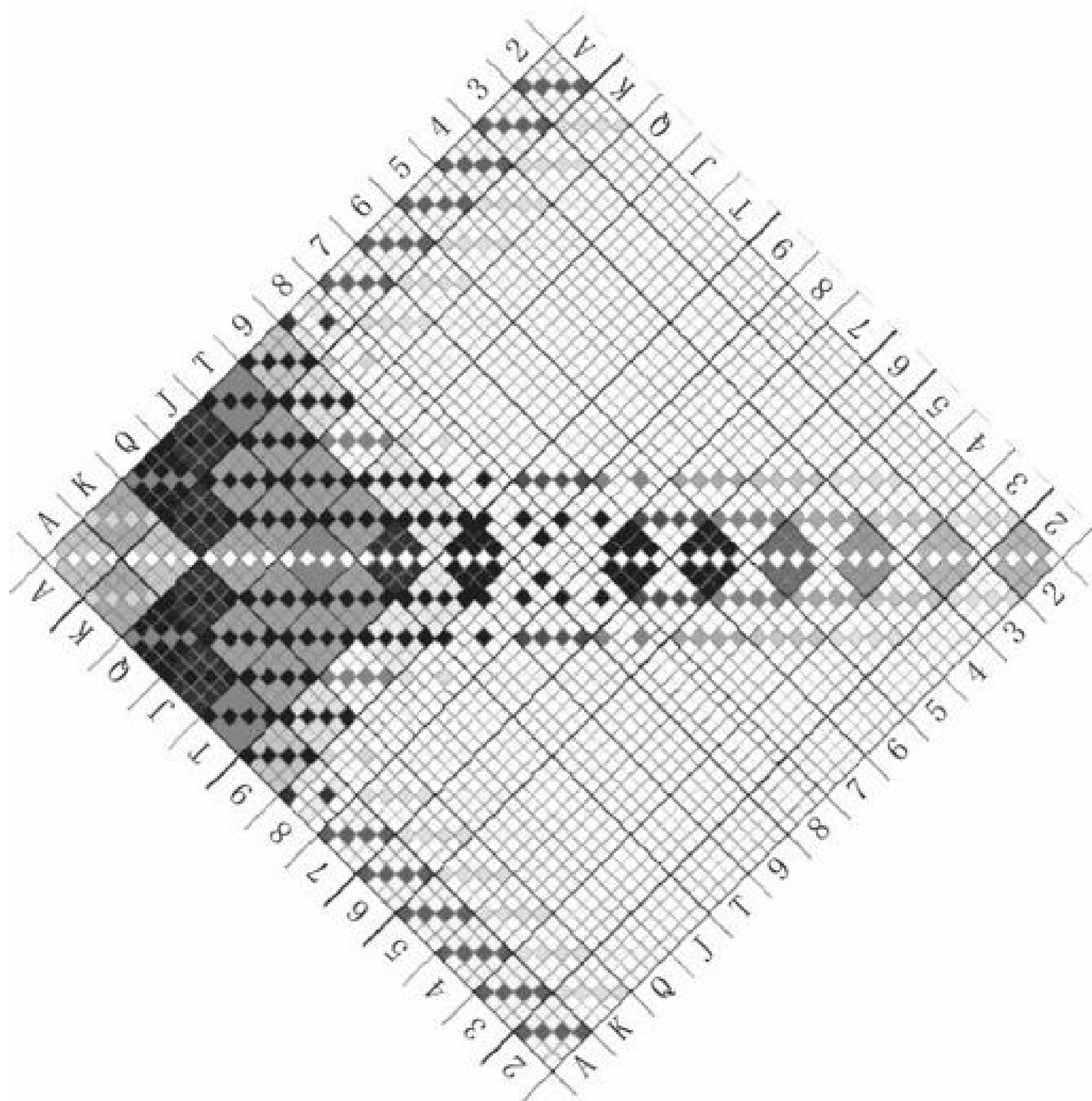


图 **10–2** 对手底牌范围的概率表达

然而，在现实情况下，如图 10–2 这般复杂的矩阵实在难以计算，玩家试图做的就是把对方的手牌范围拆分成组，每组能实现的功能是相同的（表 10–1）。这样，我们要考虑的就是“律师”手上的底牌是否大于我们手中的一对 8。

幸运的是，对方的底牌大于一对 8 的概率很低：在德州扑克中，很少有玩家的底牌是一副对子。的确，扑克牌一经发出，底牌是一对 9 或更大牌的概率仅为 3%。但我们需要根据对方的叫注对概率的预测进行更新：“律师”已经舍弃了许多无用的牌，表明他确实拥有一副大牌。鉴

于他目前的玩法，我们估计他的底牌大于一对 8 的概率已增至 6%。

表 **10–1** 对手在前 **3** 张公共牌发出之前的底牌组

底牌类型	底牌示例	跟注之前的先验概率	跟注之后的后验概率
8 到 A 的对子	J♥ J♣	3%	6%
2 到 7 的对子	6♠ 6♦	3%	12%
同花牌	A♥ 9♥	4%	15%
非同花牌	A♠ Q♥	11%	16%
同花的人头牌	Q♠ J♠	2%	10%
非同花的人头牌	K♣ T♠	5%	16%
同花琏	7♦ 6♦	2%	8%
混合牌	J♣ 8♣	70%	17%

而另外 94%的可能，是“律师”手中的底牌其实小于我们手中的底牌。问题在于，还有另外 5 张牌没有发出，而这 5 张牌很难使我们的底牌有什么起色（除非能得到整副牌中另外两张 8 中的一张），但“律师”却可以轻而易举地凑出较大的对子、顺子或是同花。

当发牌人把前 3 张公共牌分放到桌子中央时，“律师”喝了一大口咖啡。此时，公共牌包括一张梅花 K、一张梅花 3 和一张红桃 9。

K♣ 9♥ 3♣

这 3 张牌无法改善我们的底牌，只希望对“律师”而言也是如此，这样我们手中的一对 8 仍然是最大的牌。于是，我们在 65 美元的底池中投下了较为适中的 35 美元。“律师”稍微迟疑了一下，决定跟注。

进一步对其底牌范围进行精确预测后就可以看出，“律师”的决定对我们来说可不是什么好消息。按照贝叶斯定理，最关键的是要考虑条件概率。比如，如果“律师”的底牌是 K♣ J♦，而公共牌又使他凑成一对 K，那么他继续跟注的可能性会有多大？（几乎可以肯定的是，他至少已有一副对牌才会跟注，那么他还会加注吗？）如果他的底牌小于我们的一对 8，比如说是 7♥ 7♠，那么他跟注而非弃牌的可能性又有多大呢？还是那句话，如果我们的时间充裕，就能够把 1 326 种可能性一一列出，然后对我们的预测进行相应地修改（表 10–2）。

而我们的实际预测并不能如图所示那样精确。但根据“律师”的表

现，我们还是能够推断出一些有关他的底牌的广泛概率表征：“律师”的底牌能够与那 3 张公共牌构成很好的组合，他可能会组合出一对 K 甚至更大的对子，这种情况的概率占到约 30%——除非备受压力，拥有一副好的底牌是不会轻易弃牌的。另外，“律师”的底牌小于一对 K 但大于我们的一对 8 的概率约为 20%，这同样会赢了我们，但是，如果我们继续大胆下注，“律师”或许就会退缩，选择弃牌。

此外，“律师”依靠公共牌凑成顺子或同花的概率为 25%，也就是说，他现在的牌可能小于我们，但仍有机会翻身。最后一种可能性是，他的底牌可能小于我们的底牌，或许根本就是散牌，一直不弃牌只是想要唬住我们，这种情况也有 25%的概率。

以上就是可能性最大的 4 种情况，你会发现“玩转”扑克牌并非易事。有些概率告诉我们应该继续大胆下注，有些概率则提醒我们还是谨慎为好，还有另外一些概率则暗示我们要作好弃牌的准备。

就在我们犹豫不决的时候，发牌人发出一张“救命”牌，这张牌能够让我们起死回生，那就是另外两张 8 中的 8♦，这样我们就有 3 张 8 了。而“律师”打败我们的唯一方法，就是他的底牌是一对 9 或一对 K，这样就能凑成 3 张 9 或 3 张 K，如果是这样，那说明他在玩牌时只是在假意示弱以蒙骗我们。（扑克玩家称此举为“故意示弱”，与虚张声势相对。）然而，我们不应该如此被动，在对他的底牌进行分析之后，我们认为自己获胜的概率应该能达到 98%。于是，在这一轮我们押下的赌注较高：把 100 美元投进了赌金数为 135 美元的底池。

表 **10–2** 前 **3** 张公共牌发出后，对手的底牌组合

底牌类型	底牌示例	跟注之前的先验概率	跟注之后的后验概率
3 条	9♦ 9♠	2%	2%
两对	K♥ 9♥	2%	2%
K 或更大的对子	K♠ J♦	15%	28%
8 到 Q 的对子	9♦ 8♦	13%	20%
7 或更小的对子	A♠ 3♠	15%	20%
同花，没有对子	8♣ 6♣	6%	12%

顺子（不同花）	Q♦ T♠	12%	12%
其他散牌	A♥ J♣	35%	4%

“律师”又一次跟注。由于“律师”很有可能会将他较小的底牌和较小的扶持牌都弃掉，我们就能进一步缩小其底牌的范围。实际上，在 1 326 种底牌组合中，眼下最有可能出现的牌面组合已不超过 75 种。我们之前最担忧的是“律师”可能会有一对 K，不过现在看来这个担忧是多余的。我们现在所担忧的是另一个花色为梅花的牌出现，因为这样仍然能帮他凑成一副同花牌。

不过，最后一张公共牌不过是一张毫无威胁的 5♠，并没能让“律师”凑成一副同花牌：

K♣ 9♥ 3♣ 8♦ 5♠

我们将 250 美元的赌金投入赌金数为 335 美元的底池中，想着“律师”能够带着他的臭牌跟注。然而，“律师”突然精神起来，用极低的声音对发牌人说：“全进。”接着把他剩下的约 1 200 美元的筹码都推进了底池。

表 **10–3** 第四张公共牌发出后，对手的底牌组合

底牌类型	底牌示例	跟注之前的先验概率	跟注之后的后验概率
3 张 K 或 9	9♦ 9♠	2%	1%
3 张 3	3♥ 3♣	1%	1%
两对	K♣ 9♥	2%	2%
K 或更大的对子	K♥ J♠	28%	45%
3 到 Q 的对子	9♦ 7♦	40%	33%
同花，没有对子	A♣ 2♣	12%	14%
顺子（不同花）	J♦ T♠	13%	3%
其他散牌	A♠ Q♥	3%	1%

这到底是怎么回事？我们需要运用贝叶斯定理来考虑这个问题了。如果我们此前对“律师”手牌的预测失误了，那可就要与 1 200 美元失之交臂了。

我们看了看台面意识到，1 326 种可能性中确有一副底牌似乎与“律师”现在的底牌一致，那就是 7♣ 6♣，这是一副同花连牌。所以我们猜想，“律师”在前 3 张公共牌发出之前，肯定已经有了这样一副底牌。当前 3 张公共牌发出时，这副底牌已经成了一副有 4 张梅花的同花牌，这

样我们就无法打败他。发出第四张公共牌后，同花牌无法实现，他的牌反而更大了：我们得到的公共牌是 8♦，这使得我们拥有 3 张 8，但也给了“律师”更好的机会，只要他有任何一张 10 或 5，就可以凑成顺子。如果他的底牌确实是 7♣ 6♣，最后一张公共牌 5?恰好使他凑成顺子，可以打败我们的 3 张 8，这也解释了他为什么会大胆下注。

那么，我们是否应该弃牌呢？即使从未玩过扑克牌，面对这种情况也要少安毋躁，仔细考虑一下。

答案是，千万不要弃牌。实际上，对手下注越多你越应该感到高兴，因为底池里的赌金会越来越多。

贝叶斯定理为我们作了这个解答。“律师”决定“全进”的确是极其强劲的一招，这一招传达的信息比之前所有的跟进都多。但在“律师”全进之前，我们认为他的底牌中有 7♣ 6♣的概率非常低，可能只有 1%，只是 1 326 种组合中的一种。除非我们非常确信“律师”的底牌中有 7♣ 6♣，否则弃牌就是大错特错。我们的底牌大于对方底牌的概率只要达到 35%，这次跟注就准没错。

实际上，“律师”的底牌还存在另外一些可能，如一对 3 或一对 5，如果是这类底牌仍会输给我们的一对 8。“律师”的底牌也可能是 K♥ 5♥，这样他就会有一对 K 和一对 5。有些玩家或许是一对 A。根据“律师”对我们手中底牌的推测，他有可能推断出自己的底牌大于我们的底牌，这一局面即使没有好到足以让他全进，但他也许还是想从中大捞一笔。

除了顺子之外，还有其他的牌面组合可以打败我们。如果“律师”一路上只是故意示弱，他真正的底牌是一对 9 或一对 K，那么现在他已经大获全胜了。这招与虚张声势的可能性差不多。如果“律师”错失了一副同花，那么他取胜的唯一办法仅剩下虚张声势。

阿瑟·柯南·道尔曾经说过：“除去不可能的，剩下的即使再不（大）可能，那也是真相。”这句话听上去很有道理，但问题是我们如何从“不大可能”中区别出“不可能”，区别过于仔细总会惹上麻烦。目前所有对手的底牌都是“不大可能”的，而“律师”的底牌很不常见。这是一些“不大可能”与“不可能”之间的较量，也是计算与假设之间的较量，是实际计算结果与“律师”的底牌是 7♣ 6♣这一假设之间的较量。如果用计算机计算所有的可能性，我们手持最大底牌的概率约为 2/3

表 **10–4** 第五张公共牌发出后，对手的底牌组合

底牌类型	底牌示例	跟注之前的先验概率	跟注之后的后验概率
顺子	7♣ 6♣	1%	16%

3 张 K 或 9	9♦ 9♠	2%	17%
3 张 5 或 3	5♥ 5♣	2%	19%
两对	K♥ 5♥	3%	20%
K 或更大的对子	A♠ A♦	44%	15%
8 到 Q 的对子	8♥ 7♥	35%	4%
无对（纯粹唬牌）	7♠ 2♥	13%	9%

在真实游戏中，扑克牌玩家评估底牌概率的方法各不相同。技艺高超的玩家在不确定的情况下做出合理的概率判断的能力要超出其他99.9%的人。实际上，我还没有听说过有哪一款游戏或智力训练可以提高这种评估能力。然而，当我把这副底牌发布到专业扑克牌玩家共享的“二加二”在线论坛后，各种评价接连不断，有些人认为我们的手牌准是最大底牌，也有人认为我们被“律师”吃定了。我认为这两种评价都过于自信了。在对对手的底牌一无所知的情况下，我们不能贸然行动，但我们的预测错误普遍来自对确定性的错误认识，认为真实世界的确定性和我们想象的别无二致。在这种情况下，准确推断出对方的底牌就意味着弃牌，而全面评价各种可能性后——伴随着底池里的赌金不断增加——我们的结论是继续跟注。

如果这样一副底牌出现在 ESPN 播放的扑克牌锦标赛上，观众可以看到每位玩家的底牌，那么来自解说员的分析就是另一番模样了。如果观众知道对方的底牌是 7♣ 6♣的话，肯定会断言我方将弃牌。而如果我们选择弃牌后发现对方的底牌不过是 3♥ 3♠，那么他们肯定会表现出异常的惊讶，认为我们应该往底池中投下更多赌金。

在 2009 年的一场电视转播的扑克牌比赛中，比赛双方是两名世界级玩家，汤姆·德万和菲尔·艾维。比赛过程中，底池里的赌金最终超过了 100 万美元。艾维在第四张公共牌发出后，竟奇迹般地凑成一副 5 张牌的顺子，而不幸的是，同时也让德万凑成了一副 7 张牌的顺子，这是德万的牌大于艾维的唯一可能。一位解说员说道，“若有人要退出比赛，非菲尔·艾维莫属。”这句话暗指艾维选择弃牌才是明智之举。实际上，弃牌的做法是极其糟糕的。根据当时艾维对情况的了解以及双方各不相让的阵势，我们猜测他当时应该认识到自己手中的牌有 90%的概率是最大的牌。此时，如果艾维还选择弃牌，那就是牌技过差了。

电视报道对扑克牌游戏来说确实是一个福祉，但也误导了许多临时

玩家，使得他们把太多重点放在比赛的结果上，而忽略了做出正确决定的过程，并且他们还认为这就是正确的打牌方式。

德万对我说，“把对手的底牌缩小到一种可能性的情况并不常见，而电视节目却会让你相信你可以做到这一点。”

虚张声势，让对手猜不出你的底牌

德万有一个非常出名的网络昵称叫“durrrr”，之所以选了这样一个昵称，是因为他觉得这个昵称能使其他玩家输牌之后情绪失控。德万在 17 岁时就在网上的“全速扑克室”（全球性虚拟扑克室）投入了 50 美元的赌注，之后从波士顿大学辍学，成为一名全职扑克牌玩家，并在“在线扑克食物链”中一路升级为顶级掠食者。每个月他都有上百万美元进账，有时他也会输钱，但大多数时候都会赢。

2012 年我与德万有过一次对话，当时人们普遍认为他是全世界最棒的无注限德州扑克玩家之一。人们都知晓德万勇于创新、进攻主动，最重要的是毫不畏惧。2009 年，德万向全球扑克牌玩家发起挑战，进行一对一的比赛，除了他的好朋友菲尔·加尔福特之外，其他对手全都惨败。3 位扑克牌高手相继与其鏖战，德万战胜了其中的两位。

虽然打牌时比较高调，但德万为人却十分低调，总体来看，德万思考扑克牌游戏和思考世界的方式是非常符合概率方法的。他之所以能赢得比赛，是因为他的对手总是过于自信。德万对我说：“生活中的大多数领域都需要用概率的方法思考问题，而不是简单的是或否。不论是分析建立税收联盟、购买食品，还是分析自己是否会被解雇，在很多领域中，人们其实都存在巨大的缺陷。”

德万总会设法搞乱比赛、迷惑对手，以利用这些人性的缺陷。如果说扑克牌比赛中最重要的技术是学会如何预测对手的底牌，那么第二重要的就是学会反预测。德万说：“一个人的牌技越高，我们就越不能确定他的底牌是什么，越不能确定他在做什么、他的手牌范围是什么。而且，他常常会利用你对他的判断做文章。”

虽然我永远不可能成为德万那样的玩家，但作为职业扑克牌手期间，我也在以自己的方式这样做。2000 年年中，在一些不那么激烈的在线扑克牌游戏中，我打得保守且谨慎，虽然也能赢钱，但我很快发现，更为主动的方式才能赢更多的钱。于是，我就想找到对手预测我的底牌范围时可能出现的盲点。

比如，发出前 3 张公共牌时，如果你加注，那么你的对手自然会认为你的底牌非常大，其中可能有 A、K 或 Q，当然，有时的确如此。但如果你的底牌同“律师”的一样，是 7♣ 6♣这样的小牌，同样可以加注。

我的发现是，当 A 或 K 这样的大牌出现在台面上时，对手通常会相信我有可配套的牌，于是选择弃牌。而如果是一些小牌，我也总能凑成对子或其他好牌。有时，我甚至可以用这些公共牌凑成一副看似不大可能凑成的顺子，使对手变得情绪失控，开始像疯子一样乱玩。扑克牌有意思的地方在于，虽说原因各异，但牌技最佳的玩家和牌技最差的玩家玩牌时都可以随心所欲地出牌。于是，即使很有可能赢得对手的赌金，你仍然可以假装自己是个臭手，让对手放松警惕。

最终，有些对手识破了我主动出击的牌风，但也无妨，因为当我的底牌变得“可预测”时，如底牌为一对 K，对手会更愿意跟注，这样我就能够赢更多的钱。

实际上，虚张声势和主动出击的玩法在扑克牌游戏中是很常见的，而且也是很有必要的，否则对手轻而易举就能识破你的底牌。从 5 年前开始，我不再定期玩扑克牌，此后扑克牌游戏中主动出击的牌风变得极为普遍，博弈论和计算机仿真程序都强烈地指出，这是最佳玩法。利用大量的可能性闪电攻击你的对手，是使他的概率计算变得复杂的最好办法。

有时，你可能会发现对手在某些状况下对概率的直观预估过于粗糙。当一个扑克牌玩家认为对手绝不会以某种方式——比如虚张声势的方式——出牌时，这时你便有机会混淆他对“不大可能”和“不可能”的认识，以达到利用他的盲点的目的。

德万说：“我知道我做过的许多事情并不是最佳选择，但长期来看，恰恰是这些事情为我赢得了大笔赌金。但直到最近几年，人们才终于开始意识到这一点，情况才有所改善。”

无注限德州扑克是德万最在行的游戏，也最适合采用这个游戏策略，因为牌手能够通过赌注的大小控制每个决定可能输掉的赌金。德万有些选择涉及的赌金不超过 100 美元，而有些则高达 10 000 美元、100 000 美元，甚至更多。只要能在有百万美元赌金的底池中做出几个额外的正确决定，那么每次输掉的 100 美元就算累积起来也不算什么。

而我大多数时候玩的是有注限德州扑克，每轮的注额增量是固定的。（有注限德州扑克一直是扑克锦标赛以外最为流行的玩法；10 年前，在美国玩无注限德州扑克的地区通常不过两三个。）有注限德州扑克的创新性少一些。在实践赶上理论之前，我利用主动出击的玩法还是辉煌过几年的。在 2004~2005 年这两年，我从扑克牌游戏上获得的收入达到了 6 位数，加上巅峰时期的收益，累计约为 400 000 美元。

学习曲线与二八原则亦适用于预测领域

我和德万的不同之处在于，无论赌金是多少，他几乎在任何时候都愿意与任何玩家打牌，而我仅与中上阶层的玩家打牌，而且为了赢钱，少不了陪一些水平较低的玩家打牌。幸运的是，扑克牌游戏兴起的头几年，水平较低的玩家比比皆是，我们称之为“菜鸟”。

有一条学习曲线可以应用于扑克牌游戏和其他大多数涉及预测的任务。学习曲线最关键的一点，就是它确实是一条曲线：在完成任务的过程中，我们所取得的进步不是直线性的，通常如图 10–3 所示，它是一条曲线，我将其称为帕累托曲线。

你看到的是一个横轴为“努力程度”、纵轴为“准确程度”的曲线。你可以将轴线标记为任何名称，如“经验”和“技能”，但总体思想不会改变。“努力程度”或“经验”是指在预测某一个问题上投入的金钱、时间或批判性思维，而“准确程度”或“技能”是指这一个预测在真实世界的可靠程度。

该曲线的名字来自著名的商业法则，名为帕累托原理或“二八定律”（即 80%的公司利润来自 20%的重要客户）。正如我在这里所说的，把几件最基本的事情做好会有很大的帮助。在扑克牌游戏中，只要简单地学会在牌不好的时候弃牌，在牌好的时候加注，努力预测对手的底牌，就会大大地减少自己的损失。如果你真的这样做了，也许你只用了 20%的时间学习如何玩扑克牌游戏，但也许其余 80%的时间里你会和德万这样技术一流的玩家一样明智。

在其他许多倚重预测的领域中，二八法则仍然成立。第一次成功的经历通常始于正确的数据、正确的技术和正确的动机。我们需要一些信息，越多越好，并且能保证其准确性。我们还需要熟练地掌握技术，越高端越好，但更重要的是知道如何使用这些信息和技术。还要关注准确性，抓住客观真相，阿谀奉承的预测并非我们想要的，出镜成名也不是我们的目的。

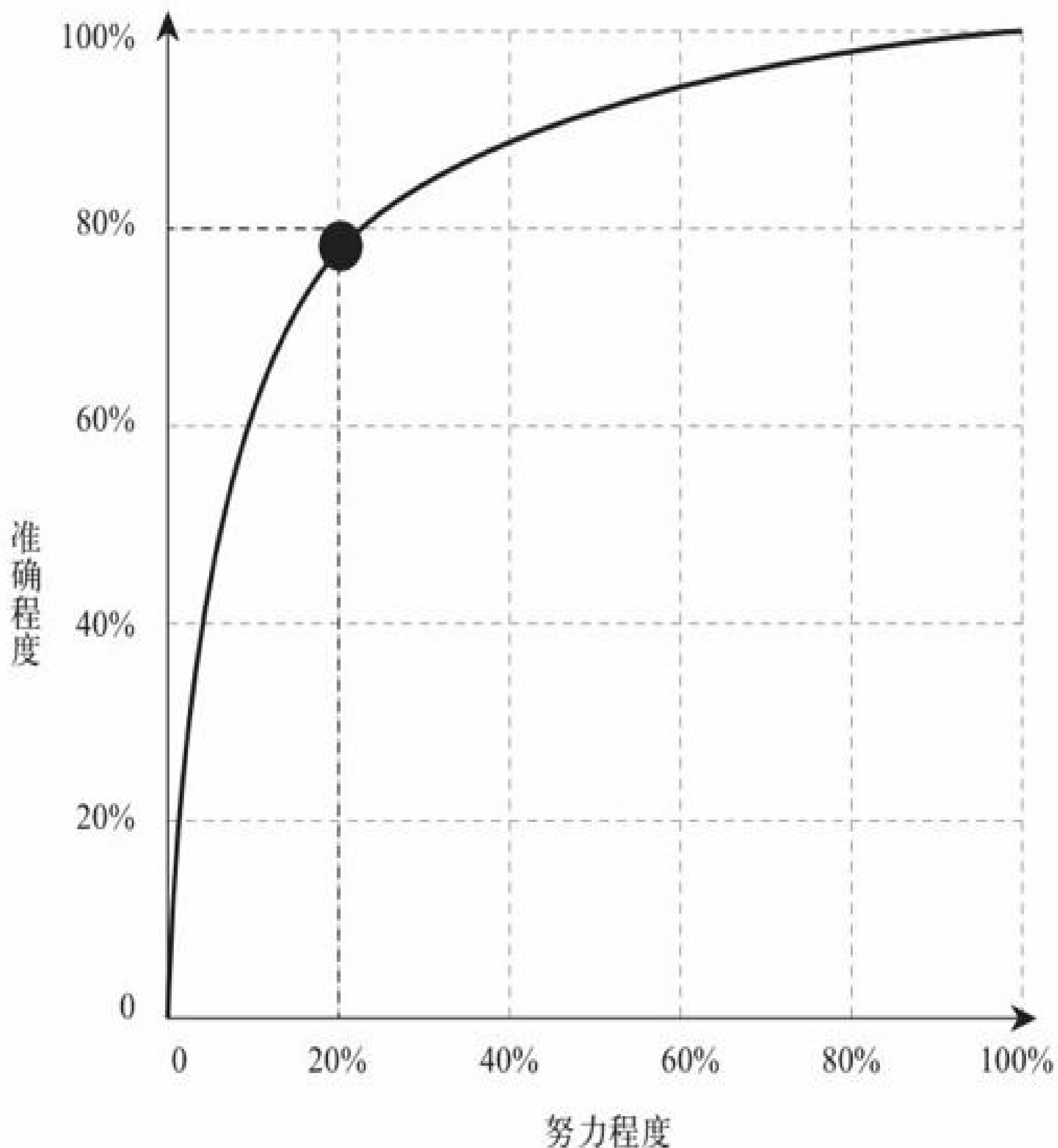


图 **10–3** 帕累托曲线

然后，就会进入中间环节，基于经验和常识以及一些系统处理过程（而非临时随意地）建立一些经验法则（启发法）。

这些并非易事，许多人都会出错。但也并不是很难，只要按照这一个方法去做，普通人也能和预测专家一样做出可信度较高的预测。

然而有些时候，重要的不是你的预测本身有多好，而是这些预测和竞争有多大关系。在扑克牌游戏中，即便你能做出的决定的正确率为

95%，但仍会输个精光，因为你的对手做出的决定的正确率为 99%。同样的，要想在股市中叱咤风云，也需要一些投资团队来为你作预测，投资团队的成员必须是精英人才，身着高档套装，拥有工商管理学硕士（MBA）学历，毕业于常春藤高校，拿着七位数的薪水，拎着高配置的手提电脑。

在这种情况下，想要打败别人，就要加倍付出努力。你会发现，自己很快就会遭遇收益递减的窘境。你收获的额外经验、策略中新添的妙计和预测模型中的附加变量，所有这些，并不能使你的牌技有所长进。同时，你建立的那些经验法则虽然有用，但现在你需要学习经验以外的更多东西。

然而，在竞争十分激烈的领域，只有提高边际效益才能挣到钱。竞争为人们设定了"水位"，而个人的利润只是"冰山一角"：漂浮在水面的仅仅是一小部分竞争优势，而隐藏在水面下支持它的，是一个由汗水铸成的巨大堡垒。

我曾经努力想避开这些领域，转而在一些"水位"很低的领域下功夫，把最基础的做好，也能得到我想要的。在"点球成金"之前的年代，棒球也是"低水位"领域之一，所以当时比利·比恩仅凭借对一些小事——如上垒率比击球率能更好地衡量一个球员的进攻表现——的认识就赚得盆满钵满。现在，很多人都意识到了这一点。在政治领域，如果再多几个"538"这样的网站，我就知道自己没多大赚头了。但通常情况下，我足以同政治评论员"竞争"，比如同"麦克劳夫伦讨论小组"的那些评论员，他们做预测时甚至连尽量准确都做不到。2000 年年中的扑克牌游戏也是如此，一些新手自以为仅从电视转播中就可以学会如何打牌，于是源源不断地涌进扑克牌领域，正是他们拉低了该领域的"水位"。

如果你具有很强的分析能力，并且能够在许多领域中发挥这种能力，那真该仔细考虑一下竞争实力。在有些领域，竞争屈服于不良动机、不良习惯以及对传统的偏执，如果你很擅长对这样的领域作预测——你有更可靠的数据或更过硬的技术——通常就有可能小赚一笔。而在人人都能对基本情况预测准确的领域中，想要脱颖而出的难度就更大了，如果这时你还认为自己拥有很大的优势，那可就是自欺欺人了。

总的来说，即使常常是事倍功半，社会也仍需要为预测付出额外的努力，或者我们应该意识到我们做的近似估计总是伴随着权衡发生的。但是，如果把预测当作商业手段，那你最好还是另谋高就吧。

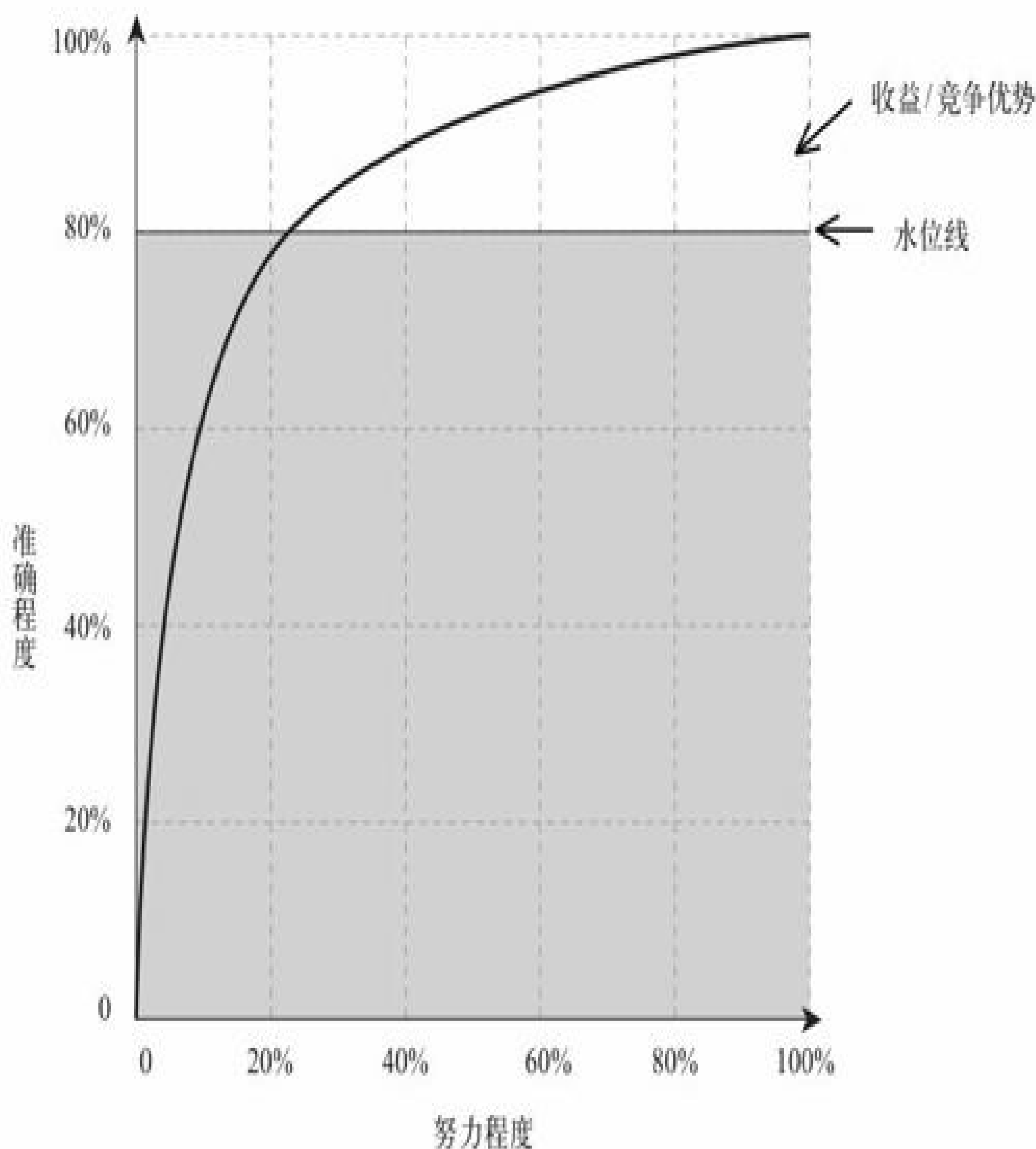


图 **10-4** 处于竞争环境下的帕累托曲线

扑克牌经济的繁荣

帕累托原理暗含的意思是，那些最差劲儿的预测者——这些人所作的预测没有一个是成功的——与最好的预测者相差太多。换句话说，水平一般的预测者更接近“水池”的顶部，而不是底部。我知道，如果我与德万打牌，肯定会输掉一大笔钱，但如果有人可以保证我还能与街上随

便找来的普通人以同样的赌金赌上一局，把我输的钱再赢回来，我便愿意跟德万赌上一局。

我们可以通过检查扑克牌玩家的统计记录来检验这一假设。我对一家在线扑克牌游戏网站的数据进行了评估，其中包括 2008~2009 年的无注限德州扑克牌玩家的随机样本。这些数据按赌金大小显示出玩家每局的输赢状况。

因为短期盈亏包含很多运气的成分，所以我使用了一个统计程序，对每个玩家的长期真实收益状况进行预测。然后，根据玩家的技能水平排序，把他们分为 10 类，第一类玩家即牌桌上排名在前 10%的玩家[3]，相当于典型的 10 人桌上的最佳玩家，而第十类则完全是“菜鸟”级玩家。

图 10–5A 显示的是，根据每一位玩家在盲注为 5 美元或 10 美元的无注限德州扑克中每 100 局的输赢情况，我对每一类玩家的真实技能都作了预测。图中包含了赢得或输给其他玩家或赌场的钱，付给赌场的钱只占每个底池（被称为抽成）或分牌所需费用（以小时计算）的一小部分。

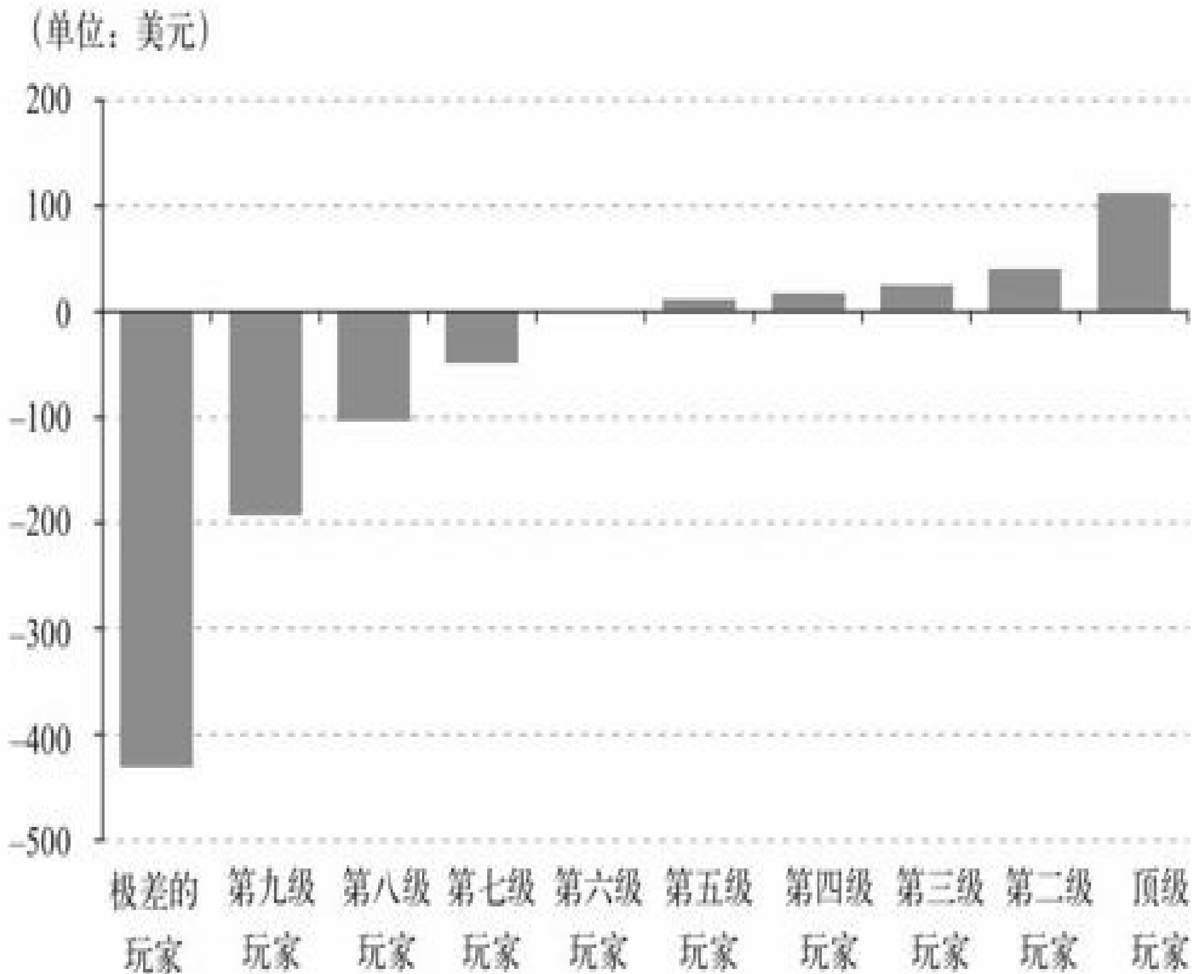


图 **10–5A** 盲注为 **5** 美元或 **10** 美元的无注限德州扑克每 **100** 局的输赢预测

据我估计，长期看来，最棒的玩家每 100 局下来平均每局的收益为 110 美元，这是在线赌场中较为客观的收入，因为那里发牌极快，仅一两个小时内几乎就能玩 100 局左右。传统的赌场之所以不那么吸引人，就是因为玩同样多的局数可能要花费 4 个小时，这样每小时只能有 25~30 美元的收益。

然而，关键的问题是，低级玩家输钱的速度甚至超过了那些高级玩家赢钱的速度。比如，我预测到最差的玩家，也就是“菜鸟”玩家，每 100 局会输掉至少 400 美元。这些玩家牌技太差了，对于他们来说弃牌也许是最佳选择，选择弃牌每 100 局只会输掉 150 美元。

由此，你可以看到二八法则的统计学回声：低级玩家与一般水平玩家之间的差异比一般水平玩家与高级玩家之间的差异大得多。技能较好的玩家彼此之间所差无几，而技能较差的玩家甚至会犯最低级的错误，通常与最佳策略背道而驰。

在经典的关于扑克牌游戏的电影《赌王之王》中，马特·达蒙饰演的角色告诉我们，如果你在半个小时内无法打败弱者，你就会被弱者打败。我觉得这句话不一定正确：也许比赛中并没有弱者。然而，如果在比赛中无法打败一两个差劲的对手，那么很明显，你可能不应该参与其中。扑克牌游戏输赢难料，即使是一个“菜鸟”级玩家也可能会大有作为。

在我刚刚描述的那场比赛中，一个“菜鸟”级玩家有几个月没碰过扑克牌，他的出现对其他玩家来说，每 100 局能收获 40 美元，扣除牌桌费用，足以让牌桌上一半的人赢钱。扑克牌游戏遵循着与“涓滴效应”相反的财富“逆流”理论：技术最差的 10%的玩家输钱的速度足以使大量中等水平玩家的输赢相抵。

但是如果那些“菜鸟”级玩家，也就是那些传说中的弱者被踢出局了呢？要知道，输钱快的人注定会被踢出局，这使得一些本来赢钱的人反而会成为输钱的人（图 10–5B）。现在，我们可以预测，只有那些顶级玩家才能长期稳赚不赔，但是赢钱的速度会比前几局稍慢一些。

另外，“菜鸟”级玩家数量的减少对其他玩家来说，会产生一种“连带效应”。比之前“菜鸟”级玩家技术稍好一些的玩家此时成了牌技最差的玩家，输钱的速度比之前要快。所以，这些玩家也会被踢出局，而这对于剩下的玩家来说将是更大的挑战。扑克牌生态系统中的整个平衡被打破了。

如果低水平的玩家不断地被踢出局，扑克牌游戏该如何维持呢？有时，“菜鸟”级玩家实际上输得很不值得：PokerKingBlog.com 网站爆料，那个叫盖·拉利伯特的人其实是太阳剧团的首席执行官，他在 2008 年的在线扑克牌游戏中输掉了近 1 700 万美元，他还想以最高的赌金与德万一决高下。不论赌金是多少，作为亿万富翁的拉利伯德其实只是为了挑战智力而玩牌，输钱对他来说根本不算什么，就像普通人在“21 点”赌博中输了几百美元一样。

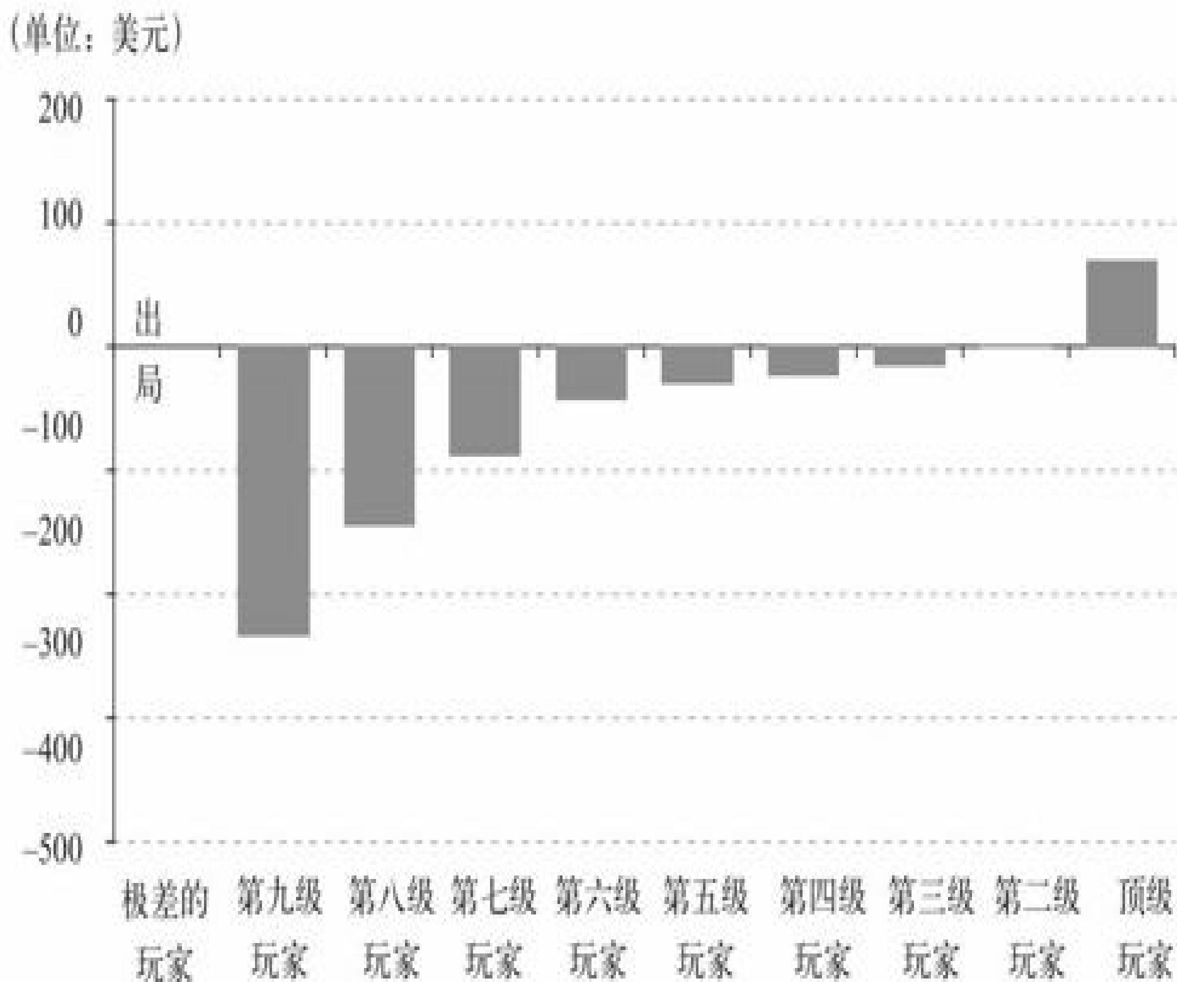


图 **10–5B** “菜鸟”级玩家出局后，盲注为 **5** 美元或 **10** 美元的无注限德州扑克每 **100** 局的输赢预测

更为常见的答案是，输钱的“菜鸟”级玩家永远不止一个，每个牌局中“菜鸟”级玩家的数量相对稳定，他们输上几百或几千美元，然后被踢出局。在百乐宫这样的真实赌场中，这些玩家可能是在双骰桌或在百乐宫旗下某家夜总会试过身手之后来到这里的，又或者是在某个联赛或是小赌注比赛中连赢几局之后来到这里的。

在我经历的在线扑克牌游戏中，“菜鸟”级玩家的人数并不稳定，其数量取决于不同国家的监管环境、扑克牌网站的广告宣传力度以及每年的游戏开始时间。在扑克牌游戏刚刚兴起的那几年，玩家人数急剧上升，其中不乏大量的“菜鸟”级玩家。

但是，这种繁荣的现象很快就发生了改变。

扑克牌经济泡沫的破灭

2006 年 10 月，即将离任的美国共和党议会想要在美国中期选举之前赢得“有价值选民”的支持，但又受到一些紧急问题的限制，于是通过了一项含糊不清的“非法互联网赌博强制法案”（UIGEA）。严格地说，这一项法案并没有指出在线扑克牌游戏是非法的，它只是把目标指向了操纵金钱进出扑克牌游戏网站的第三方。该法案规定，人们可以在线玩牌，但不可以涉赌。同时，美国司法部也把矛头指向那些为美国人提供在线赌博服务的公司。美国境外网站 BetOnSports 公共有限公司的首席执行官戴维·卡罗瑟斯在达拉斯转乘从英国飞往哥斯达黎加的飞机时被逮捕。其他行动也随之而来。

所有这些都使在线扑克牌游戏玩家还有网站经营者闻风丧胆，当时最大的在线扑克牌游戏网站 Party Poker，在“非法互联网赌博强制法案”颁布后停业两周，其股票也在 24 个小时内暴跌 65%。其他网站则是上有政策、下有对策，照常经营，但再想从玩家腰包里拿钱就阻力重重了。

Party Poker 网站在广告宣传上下了很大的功夫，而且以拥有技术最差的“菜鸟级”玩家而闻名，我玩牌的大部分钱都是从这里赢的。这家网站在宣布停业的两周内仍向美国人开放，但开放的都是极其简单的游戏，有时甚至只能玩《蝇王》这样的儿童智力游戏。在那段时间，我的扑克牌生涯达到了赢钱的顶峰。

然而，在 Party Poker 网站将美国人拒之门外之后，我就“转战”到一些更为刺激的网站，如 PokerStars 网站，在那里，我鲜有收获。实际上那时我已经开始输钱了，而且输了很多钱：在 2006 年最后的几个月里，我输掉了约 75 000 美元，其中大部分的钱都是在一个恐怖的夜晚输掉的。2007 年年初的几个月，我继续输钱的走势，又输了大概 60 000 美元。至此，我再也没有翻盘的信心了，我把剩余的钱兑现，并决定退出这个游戏。

当时我得出的结论是，如今的扑克牌玩家群体的构成已经与从前大不相同了。一些以此为生的职业玩家还在继续坚守，而大多数业余玩家或赌光积蓄或直接破产，不得不退出。扑克牌经济脆弱的生态系统被打破了，没有那些低级玩家的支撑，这个领域的“水位”已经上升，老手也变成了新手。

与此同时，我的牌技在“非法互联网赌博强制法案”颁布之前其实就已经开始倒退，至少是停滞不前。过去玩牌时，我总是把职业玩家拥有的那种危险特质——认为我有资格赢钱——和低级玩家的坏习惯结合在一起，赌博到深夜，即使已经和朋友玩到很晚，也从不耽误玩牌。我这种无创新、无灵感的玩法已经到极限了。

回顾过去，我还是很幸运的。利用业余时间——随着“非法互联网赌博强制法案”的颁布，我对政治产生了日益浓厚的兴趣——我最终建成了“538”网站。尽管将赢来的钱输掉了 1/3 让我感觉不舒服，但总好过把钱全部输光，而那些继续玩牌的玩家的运气就没有这么好了。2011 年，美国司法部提交了诉状——永久关闭在线扑克牌网站，这一天被在线扑克牌网站称为“黑色星期五”，其中一些网站被证明无偿还能力，无法退还玩家的赌金。

我有时会想，如果这事发生在我身上会怎么样。一个理论上的长胜玩家确实有可能连续数月或一整年都只输不赢。一个总输钱的玩家也有可能连赢几局之后，才意识到自己水平不够。扑克牌游戏就是如此捉摸不定，难以掌控。

既靠运气也拼技能

运气和技能通常被视为两个极端，但两者之间的关系其实更复杂一些。

比方说，我们都认为美国职业棒球大联盟中的队员都是拥有高技能的职业球员。想以时速 158 公里的速度把棒球击打出去并非易事，而有些人确实比别人更具有这方面的天赋。但是，棒球比赛中也有运气成分，击球手可以玩命地击球，却仍会被二垒手接杀，这就是运气不好的表现。技能上的区别需要很久才能弄明白，甚至用上几个月的统计数据分析恐怕都不够。在图 10–6 中，我在横轴上标出了 2011 年 4 月美国职业棒球大联盟球员的平均击球率，纵轴是 2011 年 5 月的情况，两者之间似乎并无关系，（比如，一位名叫布伦达·瑞安的球员 4 月的击球数为 184 次，5 月的击球数则为 384 次。）但是，从长期的统计数据来看，在棒球队员的整个赛季或是整个职业生涯中，击球的能力确实因人而异。

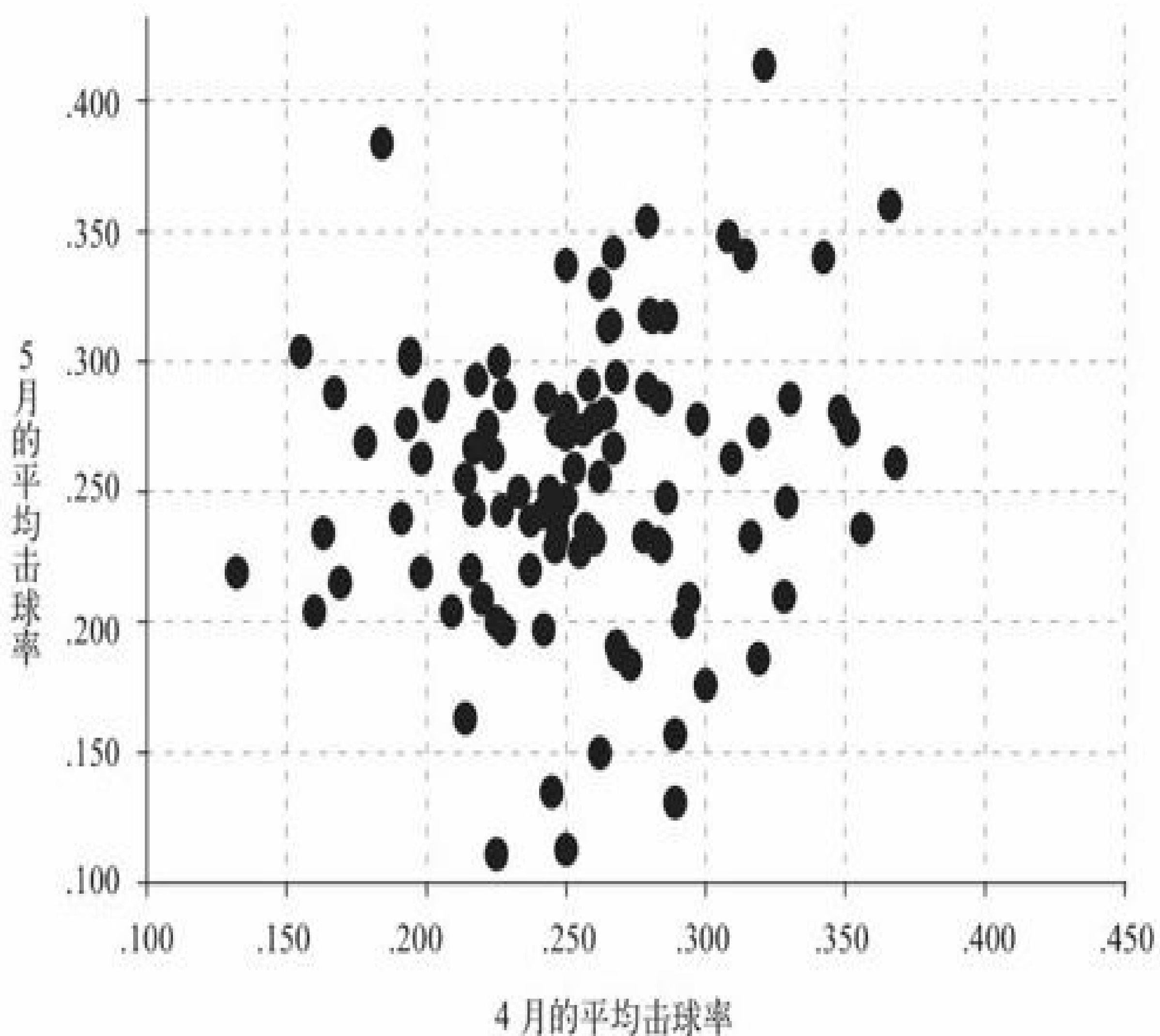


图 **10–6 2011** 年 **4** 月与 **5** 月，美国职业棒球大联盟球员的平均击球率

在运气和技能方面，扑克牌和棒球有些相似，扑克牌游戏也是运气和技能并存。与扑克牌游戏截然不同的是“一字棋”之类的游戏（表 10–5），后者中没有运气成分，也没有技能可言。一个二级水平的老手就可以玩得和比尔·盖茨同样好。

表 **10–5** 技能 **VS** 运气

	运气差	运气好
技能低	一字棋	轮盘赌
技能高	国际象棋	扑克牌

当然，扑克牌玩家需要很长时间才能明白自己的真实水平。在我曾经专攻的德州扑克中，运气成分对我的影响更大。在这个游戏中，拥有

正确的策略就意味着你将为底池中的赌金而搏杀一把，若手中的牌一直保留到最后，那就意味着这一手牌是在拼运气了。一个技能高超的德州扑克玩家在玩盲注为 100 美元和 200 美元的游戏时，每 100 局就能获益 200 美元。然而，正如名为“标准偏差”的统计学方法所测出的那样，技能高的玩家的收益也会出现波动，极有可能会达到 200 美元的 16 倍，即每 100 局的收益为 3 200 美元。

这就意味着，几万局游戏之后，技能高的玩家也有可能会落后，技能低的玩家也有可能会反超。在图 10–7 中，我用刚才提到的数据模拟了某一个玩家可能的收益和损失统计图，阴影部分表示该玩家输赢的范围，足以覆盖 95%的可能性。60 000 局之后——如果这个玩家全年每周用在赌场的时间长达 40 个小时，就能玩够 60 000 局——该玩家或者能够赢得 275 000 美元，或者会输掉 35 000 美元，后一种情况在本质上相当于这位玩家在这一年里每天都去工作，可仍然在赔钱。这就是为什么人们总说扑克牌游戏是一个难以维持生计的职业。

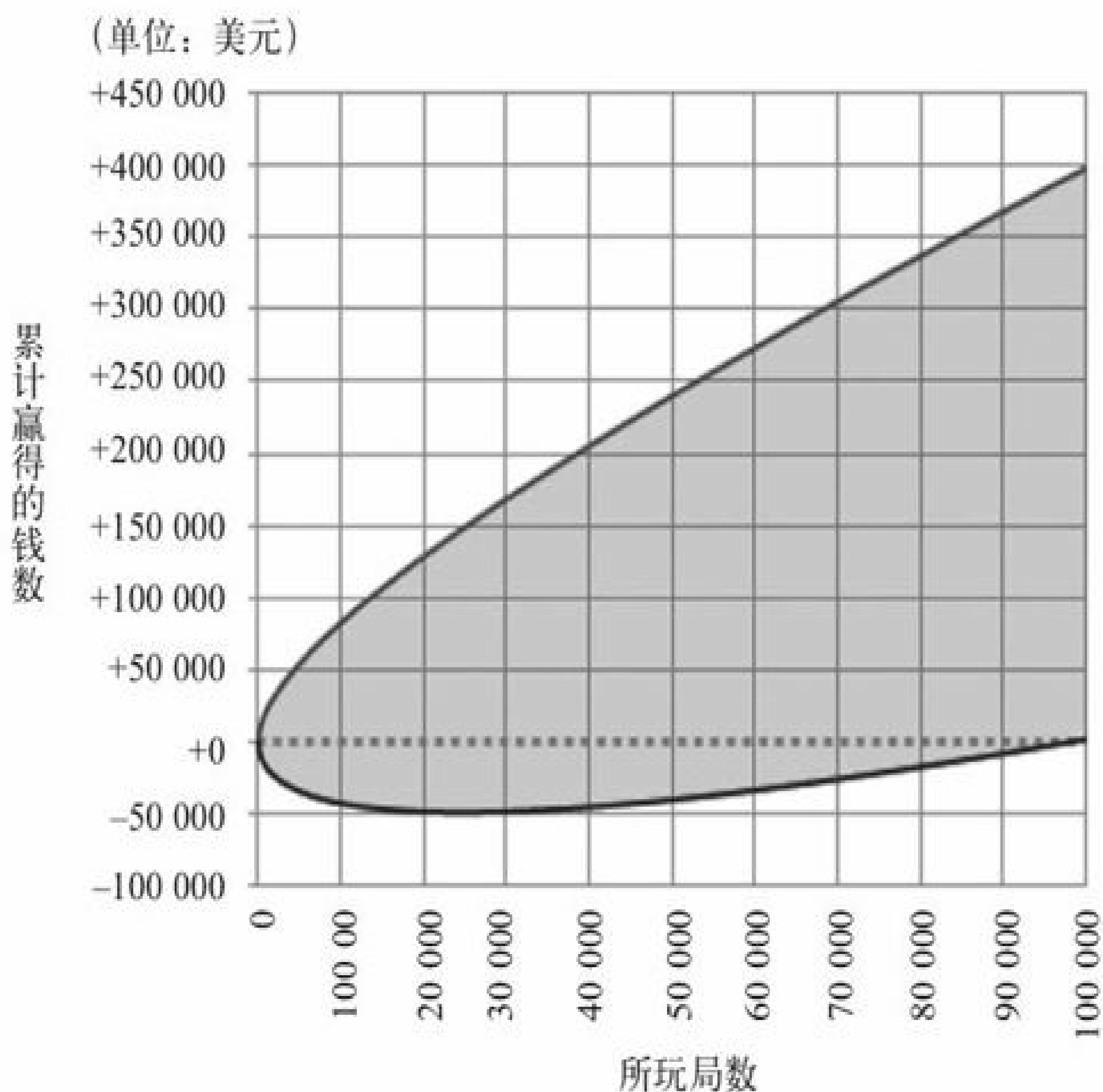


图 **10–7** 在盲注为 **100** 美元或 **200** 美元的有注限德州扑克中，技能型玩家的损益

当然，如果这位玩家通过某些途径知道自己会成为一个长期的赢家，那他就有理由挨过不停输钱的阶段。而在现实生活中，这位玩家无从得知这一点，除非是运用贝叶斯定理，根据每局的结果和之前的预测重新审视自己的牌技才行。

如果这位玩家足够诚实，那么，即使曾经辉煌过，他也应以怀疑的态度看待自己的成功。他应该认识到，只要是技能中等的玩家都会输钱，因为庄家需要资金维持扑克牌游戏的运作，而剩下的那部分钱才会在玩家之间流通。《扑克中的数学》一书中提到了这样一个观点，在盲注为 100 美元或 200 美元的有注限德州扑克游戏中，一个在前 10 000

局中赢得30 000美元的玩家，其实更有可能是一个长期的输家。

我们对于扑克牌游戏的妄想

相信你也猜到了，大多数玩家都不能如此诚实地看待自己，我本人在扑克牌经济泡沫时期当然也没有做到这一点。相反的，那些玩家一开始就认为自己会成为常胜将军，直到他们被现实击败（只有这样，他们才会知道自己也会输钱）。

德万对我说："扑克牌就是这么让人自命不凡，而人们玩扑克牌时总是想当然。"

另一位玩家达尔斯·比林斯创建了一套计算机程序，在与全球部分顶级德州扑克玩家成功对抗后，他的话说得更为直白：

> 再没有什么别的游戏能让人们如此自以为是地认为自己简直就是魔术师，可实际上，他们的表现却那么差。从根本上说，这是因为人们无知地认为自己是神一般的人物，但事实恰恰相反。如果计算机程序以人类的骄傲自大为能量的源泉，那它在扑克牌游戏中将战无不胜。

当然，不单单在扑克牌游戏中存在这样的问题，我们在本书中会看到，这些批评同样适用于华尔街的交易者，这些交易者总认为自己可以击败标准普尔500指数这类市场基准，而通常情况下，他们根本做不到。从更广义的角度来说，过度自信在任何预测领域都是一个大问题。

扑克牌与轮盘赌不同，轮盘赌单纯依靠运气，如果轮盘不停地转，就没人会赢钱了。扑克牌玩家也不同于轮盘赌玩家，实际上，扑克牌玩家更像是投资者。一项调查显示，全球扑克牌玩家中拥有学士学位的人数比例是52%，是全美拥有学士学位人数比例的两倍，是彩票购买者中拥有学士学位的人数比例的4倍。大多数扑克牌玩家都明白，有些长期从事扑克牌游戏的玩家确实能从中挣钱，正是这一点使他们陷入麻烦。

汤米·安其罗在扑克牌梦冷却之前，是一个执著的追梦人。1990年，32岁的他辞去了某个乡村音乐乐队鼓手和钢琴师的工作，成为一名全职扑克牌玩家。

2012年的一次交谈中，安其罗对我说："我对扑克牌着迷，当我第一次听到职业扑克牌手这个词时，就爱上了它。'不用工作'这个想法简直太棒了，就像是我打败了社会，仅靠自己的智慧就可以赚钱一样。当时我真的想象不出还有什么事情比这个更让人兴奋了。"

但安其罗与其他玩家一样，也有大起大落的时候——不只是输赢结

果会使人大起大落，每次玩牌时的发挥也是一样。发挥好的时候，安其罗的情绪就会高涨，但不幸的是，他并不总能发挥得很好，于是他时常会处于情绪失控中。在其著作《扑克元素》一书中，安其罗回忆道："我是一个极容易情绪失控的人。"这本书记录了安其罗因洞察力不足而出现的进攻过于主动的玩牌状态。"造成我情绪失控的原因有很多：过于放松、过于紧张、过于主动、过于被动、打得激动、打得时间过长、过于劳累、潮湿闷热、优越感、恼羞烦躁、遭受不公平、沮丧、懒散、为了报复、资金不足、资金过剩、遭受羞辱、分神、惊慌、嫉妒、'比萨不合胃口'、'我被别人唬住了'，当然还有最经典的'扯平了'和'我有的是时间输钱'，也是最具'破坏'性的原因。"

安其罗最终意识到一点，尽管牌技不错，但他的情绪总会周期性失控，这个问题使得他的钱只够勉强度日。正如我们看到的那样，在扑克牌游戏中，技能低是很容易输钱的，这一点比技能高就能赢钱要容易得多。同时，即使是长期赢家，优势也并不明显。如果一个玩家在 90%的时间里处于世界顶级水平，可只要有 10%的时间出现情绪失控，那么对他来说输钱就是常态。

安其罗到了 40 岁时才全面地认识到自己情绪失控的问题，那时候他已经开始写作，并指导其他玩家玩牌。安其罗是一个洞察力很强的人，于是战略会议总会变成情绪治疗会议。

安其罗对我说："我用玩扑克牌时会出现的各种问题教导不同的人，从其他人身上很容易就能看到自己的问题。比如，有个人的智商跟我一样高，我确切地知道他对自己的牌技总是心怀幻想，于是我意识到如果每个人都有这种幻想，那我本人也不例外。"

安其罗认为，每个扑克牌玩家都会出现一定程度的情绪失控，"如果有人跑来对我说，'我不会失控，我能控制好自己的情绪'，那一定是另外一种妄想状态。这种情形时常会出现。"

玩牌时，我也有自己的"心魔"，我不是那种急得到处扔东西的人，也不会疯狂到见牌就玩（尽管相对于我那非常一般的牌技而言，我已经足够疯狂了）。有时，我也会加强自己的牌技，但是我的打法很机械，而且考虑不周，总是打持久战，通常会拖到深夜，因为我总是跟注，总是希望底池的赌金都能被我赢走。但是在内心深处，我其实早有放弃拿下比赛的想法了。

虽然在玩牌时我并不确定自己为何会情绪失控，但现在我认识到这一点了。其实最大的原因就是优越感，认为赢钱是理所当然的事情。在持久战中，当我得不到好牌而不得不弃牌时，我并不介意，因为我知道这是游戏中统计数据多样化的一部分。但我却认为自己确实打得很好，

比如，我能准确地看出对手正在虚张声势，但对手却在发出第五张公共牌之后将我打败了，这一点着实令我抓狂。我本以为自己会赢，没想到钱都被他赢走了。

情绪失控会让事物进入扭曲的状态：因为一开始就打得不好，所以输钱是必然的。扑克牌玩家会出现情绪失控的情况，最根本的原因就在于一种平衡被打破了：从短期（或通常是中长期）来看，一个玩家的输赢与其牌技高低的关系不大。玩家通常对自己的牌技没有一个实际的认识，这当然会影响游戏结果。安其罗说："我们总想紧紧地抓住能够证明我们自身理论的数据，而我们的理论通常就是'我比别人打得好'。"

以过程而不是结果为导向

在美国，我们生活在一个以结果为导向的社会中。如果某人富有、出名或是美丽，我们就会倾向于认为这些东西都是他们应得的。而通常这些事物其实都是自我强化的结果：越富有的人越能做到钱生钱，越出名的人获得的关注越多，就连美国人审美的标准也是依好莱坞明星的标准定的。

我说这番话并非有意陈述政治主张或制造政治争议，也无意支持（或反对）加强财富再分配的举措。然而，作为一个经验问题，成功取决于辛勤的汗水、自然禀赋、个人机遇和环境，换句话说，就是噪声和信号的结合。美国人更多时候倾向于强调信号的成分，除非说到自己的弱点，那时我们就会归因于运气不好了。我们以邻居房子的面积衡量其成功程度，但我们却看不到别人是经历了多少挫折才走到今天这一步的。

说到预测问题，我们真的是以结果为导向的。能够预测股市见底的投资者，即使只是借助一些古怪的统计模型恰好准确地做出了预测，也总会被冠以"天才"的名号。能够组队赢得世界职业棒球大赛冠军的经理人——即使在你查看了他的记录后发现，这支团队赢得比赛完全是因为出色的发挥，与这位经理人根本毫无关系——在人们眼里也总是比同行优秀。扑克牌更是如此，如果不是凭借营销策略的推广，克里斯·"福星"不过就是一个"赢了一大笔钱的懒汉赌徒"而已。

有时，我们总是找借口说预测不准是因为运气不好，这都是因为我们太过依赖运气了。信用评级机构在无力预测金融危机的时候也在拿运气当借口。但我们似乎也默认了一点，即当我们作预测时，总认为信号比想象得还多，在评估预测时，我们还会把准确的预测归因于拥有更多技能，而实际上并不是。

说到解决方案，其中一个就是在评价预测时要更加严格一些。一个

预测的技术含量有多高通常可以通过经验法检验出来。在有些领域中，这个目标很快就能实现，有些领域则不行。而另一个解决方案——也是数据中充满噪声时的唯一解决方案——是把重点放在过程而不是结果上。如果预测样本过于嘈杂，无法确定它是否准确，就应该查看预测者的长期预测记录，通过其预测的态度和能力来作判断。（从某种意义上讲，我们这是对预测者的预测结果进行评估。）

扑克牌玩家比别人更能理解这种说法，因为他们切身体验过大起大落的感受。像德万这样总是玩赌注很高游戏的人，可能会在一场扑克牌游戏中体验许多的不确定性，与股票投资者一生体验到的差不多。每个扑克牌玩家都多次体验过这种感觉：打得好会赢钱，也有可能输钱，打得差会输钱，也有可能赢钱。因为这些玩家知道过程与结果之间的不同。

如果你与顶级玩家有过接触，就会发现他们从不把成功看作理所当然的事情，他们一直非常重视自我提高。德万告诉我："任何对自己满意、认为自己对扑克牌比赛得心应手的人，就等着走下坡路吧。"

安其罗试图让他的客户自我提升得更快，他说："我们总是在噪声中原地打转，无法准确地看到发生了什么。"安其罗使用的方法多种多样，有时还会冲破传统，比如，他提倡冥想的方法。不是所有客户都会按他的方法做，但他的方法中的确包含着更广阔的思想，即提高学员的自我意识水平，会鼓励他们养成更好的认识能力，分清什么是可以控制的，什么是超出控制范围的。

在打牌时，我们能够控制的是打牌的过程，但我们无法控制自己拿到什么牌。如果你正确地识破了对手在虚张声势，但他仍拿到幸运牌而赢了钱，此时你应该高兴而不是生气，因为你已经尽力了。这么说是有些讽刺，但确实当你不那么关注结果时，反而会做得更好。

在这个不确定的世界上，我们仍是不完美的生物。即使作了糟糕的预测，我们也永远不会知道这是自己的错误还是预测模型的缺陷，或者只是因为我们不够走运。最接近的近似解，就是达到一种噪声与信号的和谐状态，两者缺一不可，我们要学会欣赏它们。

[1]5 美元和 10 美元的增额指的是强制赌注，称为盲注。玩家按顺时针方向下注、表态。无注限是指玩家在每轮下注过程中，可以随意在前面玩家的筹码上加注，没有任何限制。然而，在开始出牌之前，无注限的德州扑克中的赌注受到两名玩家中下注较小的玩家所下赌注的限制。如果你的对手是拥有 500 000 000 美元赌金的比尔 · 盖茨，你手中的筹码是 500 美元，这也是你能给出的最大筹码，那么盖茨给出的任何高于这一筹码的赌注都是无效的。所以，他无法强迫你去凑那 499 999 500 美元的筹码，因为你不可能仅仅为了让他摊出底牌而费尽力气去凑这笔额外的赌金。

[2]有时玩家在拥有一对 A 的情况下仍有意选择跟注而不是加注，其目的是为了蒙蔽对手。

[3]我对玩家的分析是由每位玩家分得的牌数构成的。在扑克牌游戏中，大量的牌分发给少

量的玩家，他们每天都在玩牌，而不是每月或每年才玩一次。实际上，在线扑克是“二八定律”的夸大版：来自数据库的 80% 的牌分发给 2% 的玩家。相比于那些甚至只玩过一次扑克牌的随机抽选的玩家，你遇到这 2% 的玩家的可能性更大一些，所以如果不是依据分得的牌数来分析，我的分析可能会让你对扑克牌经济产生一种不切实际的印象。

第十一章

股票市场：非理性交易者的存在让价格泡沫不可避免

2009 年，也就是 2008 年金融危机摧毁了全球经济的一年之后，纽约证券交易所一开市，美国投资者每秒的股票交易额就高达 800 万美元。一个完整的交易日结束后，交易总额达到 1 850 亿美元，大约相当于尼日利亚、菲律宾和爱尔兰等国一年的经济总量。2009 年全年，美国股票的交易总额超过 46 万亿美元，这个数字是世界 500 强企业年收入总和的 5 倍。

如此迅速的交易速度确实罕见。20 世纪 50 年代，美国一般的公司股票平均要持有 6 年后才能进行交易，这倒是与“股票是一项长期投资”的说法相吻合。但是到了 21 世纪，同样一只股票的持股期不用 6 年，而是买入 6 个月后就可进行交易,股票的交易速度提高了 12 倍，且这一速度的增长态势完全没有减缓的迹象：股票市场的交易总额经过 4~5 年就会翻一番。随着高频率交易的出现，现在纽约证券交易所中的某些股票在一微秒内就可以完成交易，速度惊人。

经济学入门课教导我们，交易双方实现共赢时，这笔交易才算是理性交易。比如，棒球队 A 拥有两名出色的游击手，但缺投手；棒球队 B 拥有很多名投手，但游击手的击球率只有 0.190，棒球队 A 用其中一名游击手交换棒球队 B 的一名投手，这算一笔理性交易。又比如，一位准备退出股市的投资者为了套现，将股票卖给一位刚入市的投资者，这也算一笔理性交易。

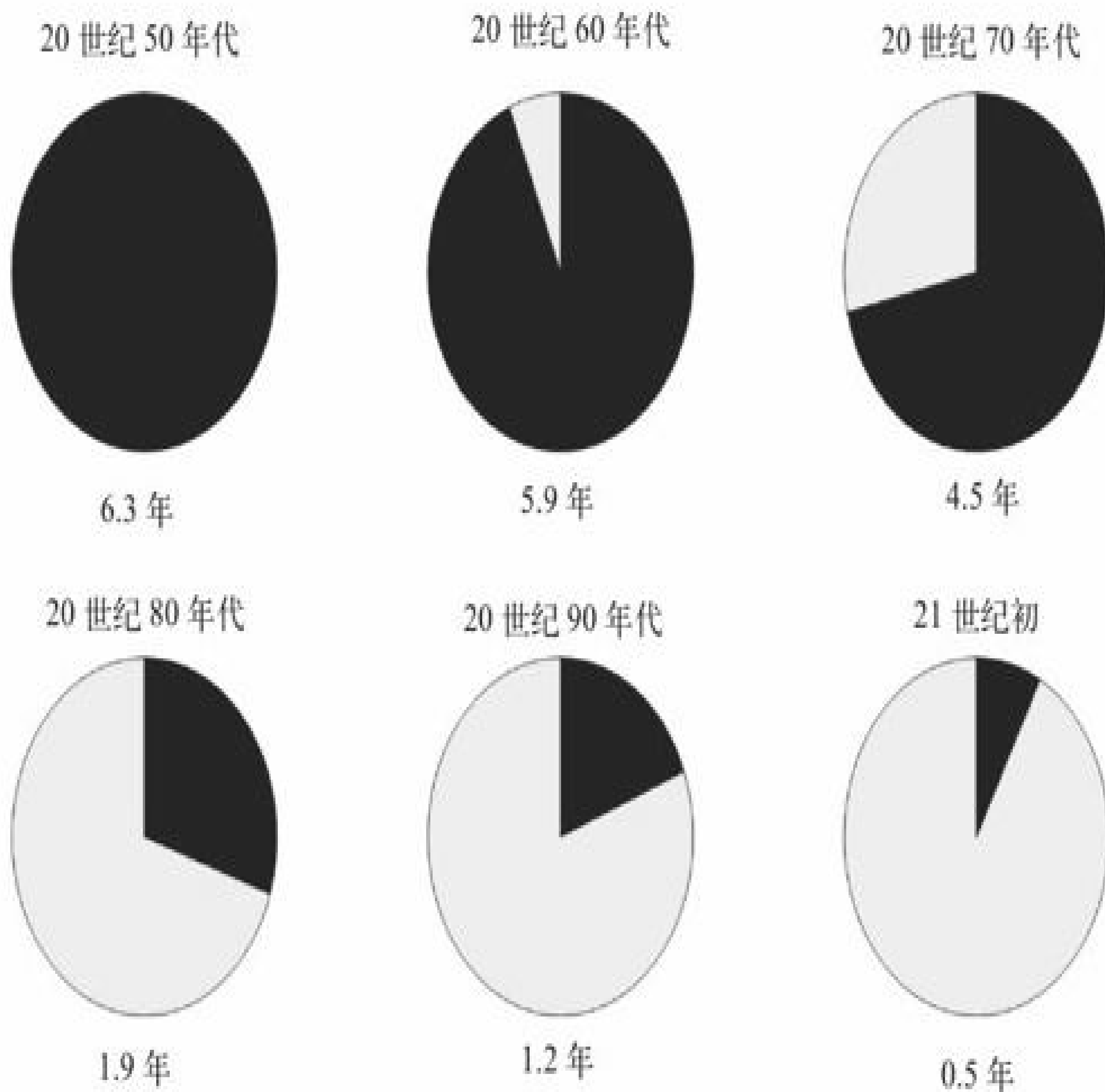


图 **11–1** 不同时期，美国一只股票需要持有的平均时长

而如今华尔街上的交易基本上都不符合这一标准，大多数交易反映的是人们对某只股票未来收益的不同看法，这些看法就是差异巨大的预测。在人类历史上，如此迅速地做出这么多高风险的预测的确是前所未有的。

为何会发生这么多笔交易，这个问题至今仍是金融界的一大谜团。越来越多的人似乎认为，他们在预测方面拥有能够胜过市场的群体智慧。那么，这些交易者理性吗？如果不是，我们能期待市场给出一个合理定价吗？

贝叶斯定理世界中的价码牌

如果你按照本书的建议遵从贝叶斯定理的指导，你就会采用一套概率性的方法来思考和预测未来。巴拉克·奥巴马连任的概率有多大？问题女星琳赛·洛翰再次被捕的概率有多大？在其他星球上发现生命的概率有多大？拉菲尔·纳达尔赢得温布尔登网球公开赛的概率有多大？一些贝叶斯论者声称，探究这些概率最明智的方法就是划定盘分线。如果你将这一个想法发挥到极致，那么在贝叶斯王国中，我们会带着很多大块的三明治价码牌随处走动，价码牌上写着我们对上述问题的概率估值。

今日价格

- 奥巴马连任……………………………55%
- 琳赛 · 洛翰被捕……………………99%
- 股市崩盘…………………………………10%
- 在火星上探测出生命……………………2%
- 纳达尔赢得温布尔登网球公开赛……30%

图 **11-2** 贝叶斯定理世界中的三明治价码牌

在贝叶斯王国里，当两个人擦肩而过并发现彼此的预测相左时，他们必须从下面两种做法中选择一种。第一种做法是修改自己的预测，使双方达成一致。如果我的三明治价码牌显示纳达尔赢得温布尔登网球公开赛的概率为 30%，而你却认为有 50%的概率，或许我们会同时将概率改为 40%。但是，不一定非得采用中间值，如果你更了解琳赛·洛翰的绯闻，或许我会认输，承认你关于琳赛·洛翰的预测更加准确，并将我的价码改成与你的一样。无论选择哪一种做法，分别时我们心底都会记住那个相同的价码——一个修订过的价码，并且我们期望这个价码能更加准确地预测现实事件的发生概率。

但是，有时我们的意见无法统一。贝叶斯定理告诉我们，如果意见不统一，我们就必须用下注的方式解决两种预测中存在的分歧。在贝叶斯王国里，要么意见达成一致，要么下注赌一把，你必须要在两者间做出选择。否则，在贝叶斯论者眼中，你就称不上真正的理性。两人商榷之后，如果你认为你的预测优于我的，那就应该很开心地对自己的预测下注，因为你已经嗅到钱的味道了。如果你并不这么认为，那就应该采纳我的预测。

当然，整个过程效率会出奇的低。我们必须对成千上万的事件做预测，并且随时都要让数百个赌注的账面清楚。而在现实世界中，市场会执行这个功能。股市中的交易依据的是固定的价码，一个大家都认同的价格，而不需要讨价还价或是事事打赌。

预测市场中的“无形的手”

事实上，自由市场论和贝叶斯定理是由同一个知识系统演变而来的。亚当·斯密和托马斯·贝叶斯是同龄人，都在苏格兰接受的教育，都深受哲学家戴维·休谟的影响。亚当·斯密的“无形之手”可以被视为贝叶斯定理的应用过程：价格受供求关系影响而上下波动，最终实现等价交换。而贝叶斯定理的推理过程也被视为一只“无形的手”，我们在为自己争辩时，也是在潜移默化地更新和改进观点，争论无果时，就会放手赌一把自己的观点。这两种情况都是寻求共识、博采众长。

由此可以断定，让市场做预测是一个特别好的办法，也是股市的真实写照：股市中包含了对某公司未来收益和分红的一系列预测结果。我认为，多数时候这个观念都是正确的。我主张预测国内生产总值（GDP）这样的经济变量时引入博彩市场。大家都期望市场竞争能迫使人们采取实际行动，而不只是做口头文章，为准确预测提供动机，进而提高预测水平。

有效市场假说让上述观点更具说服力。该假说认为，在特定条件下，各类市场的发展是不可预测的。几十年来，该假说一直是经济学界的正统学说，但考虑到近来的房地产泡沫和市场不景气，其中一些泡沫和不景气现象似乎是可预测的，于是这个假说慢慢失去了市场。但该理论的生命力还是比想象中旺盛。

本书的核心前提就是，若想做出更准确的预测，就必须承认我们的判断是不可靠的。市场交易是集体判断的反应，从这个层面来看，这些反应也会出错。事实上，能做出完美预测的市场从理论上讲是不存在的。

如果真的存在这样一个贝叶斯王国，那么博学多闻的贾斯汀·沃尔弗斯一定愿意担任那里的警察局局长，一发现有人拒绝为各自的预测下赌注时，他就会给那人开张罚单。扎着马尾、说话语速超快的沃尔弗斯是美国最优秀的青年经济学家之一，他曾向我发出一次打赌挑战——赌一顿晚饭——因为我发表了一篇博文，认为瑞克·桑托勒姆将赢得爱荷华州的党内候选人预选，这与在线预测网站 Intrade. com 的预测相悖（与我自己的预测模型的预测结果也有出入），该网站预测米特·罗姆尼会领先。在当时的情形下，我乐意赌一把，结果证明我是对的：长达一周

的重新计票工作结束后，桑托勒姆以几十票的优势赢得了此次选举。但其他时候，我都不太愿意接受沃尔弗斯的挑战。假设你和我一样都是爱打赌的人，如果你都不愿意把钱押在某个预测上，那这个预测能有多准呢？

沃尔弗斯来自澳大利亚，在悉尼时，他为一个赌马者作数据分析，以此赚取大学学费。沃尔弗斯现居费城，任职于沃顿商学院，并为“魔鬼经济学博客”撰写博文。有一次我前往沃尔弗斯的家中拜访，他真是热情好客，从萨尔科内烘焙店预订了巨无霸和三明治来招待我和我的助手阿瑞基·米利肯，以及沃尔弗斯最具才华的学生之一——戴维·罗斯柴尔德。沃尔弗斯还亲自为我的面包抹好了黄油。

沃尔弗斯和罗斯柴尔德一直在研究 Intrade. com 网站这类预测市场，这类预测网站是现实版的贝叶斯王国，在这些网站上交易的股票代表的是对现实生活的新闻预测——涵盖范围广，从奥斯卡奖得主到以色列空袭伊朗的可能性，无一不含。政治事件的预测尤其热门。例如，希拉里·克林顿能够赢得 2008 年美国民主党候选人提名权，这个新闻预测可假设为一只股票。若该预测是正确（即希拉里获得提名）而不会出现其他结果，购买这只股票的人将得到 100 美元的分红。在最终结果出来前，购买者都可以随意买卖股票。因此，股票的市场价格代表了该预测结果发生的概率。（在交易市场中，当希拉里输了爱荷华州的党内选举时，她的“股价“为每股 18 美元；当她赢得新罕布什尔州的初选时，她的“股价”升至每股 66 美元；当奥巴马最终获选时，她的“股价”则归零了。）像这样的博彩市场，对政治进行预测的传统由来已久，可追溯到 1892 年的美国总统大选——就在离美国证券交易所不远的地方，大家忙着交易代表格罗弗·克利夫兰和本杰明·哈里森的两只股票。

让我们再回到沃尔弗斯家中，在午餐开始前的几分钟，沃尔弗斯对罗斯柴尔德说：“你应该和纳特说说你那篇关于比较研究的论文。”说完露出调皮的笑容。

罗斯柴尔德说：“我曾在学术期刊上发表了一篇论文，探究无倾向性的互联网民调，并将这份民调和 2008 年预测市场的表现做了比较，结果显示其准确性不亚于预测市场。”

“你说得太委婉了，”沃尔弗斯打断他，转过来对我说，“就是让 Intrade. com 网站和你的网站对决。”

“结果，Intrade. com 网站赢了。”罗斯柴尔德紧接着说。

罗斯柴尔德的论文发表在《公共舆论季刊》上，将我在“538”网站做出的关于 2008 年美国总统大选的预测和 Intrade. com 网站做出的预测相比较。结论是，尽管“538”网站上的预测已经相当不错了，但是 Intrade.

com 网站却更胜一筹。

群体预测往往优于个体预测

我不是很认同这篇论文的研究方法论。沃尔弗斯和罗斯柴尔德在美国总统大选后对 Intrade. com 网站做了一些调整，它因此才打败了“538”网站，否则胜者就是“538”网站了。更重要的是，“538”网站发布的预测主导了 Intrade. com 网站上股票价格的走势，这就表明 Intrade. com 网站的投注者在一定程度上剽窃了他人的预测成果。

但确实，无论是经验上还是理论上都有确切的证据表明，综合不同的预测结果、利用群体智慧能让人受益匪浅。从微观经济学到政治民调，很多学科都是取众人预测结果的平均值，这个平均值要比一个人的预测准确，通常会使预测误差降低约 15%~20%。

但在取得所有事物的预测结果平均值之前，应该明白 3 件事。

第一，虽然群体预测基本上都优于典型的个体预测，但这并不意味着群体预测就一定是准确的。例如，宏观经济的群体预测就过于粗糙，提前几个月预测经济衰退也不够准确。然而，无论怎样，群体经济预测还是优于个别经济学家的预测。

第二，有确切证据表明，各种预测被综合到一起之前应当是独立做出的，只有这样，群体智慧才能发挥作用。在真实的博彩市场中（包括股票市场），人们会对彼此的行为做出反应，也确实是这样做的。因此，当一个群体开始活跃时，群体行为就会变得更加复杂。

第三，尽管群体预测优于个体预测，但这并不意味着它一定会优于最好的个体预测。例如，有一家民调公司的调查结果十分准确，所以最好采用该公司的民调结果，并且只采用这家公司的民调结果，而不是将它与其他民调公司不够准确的调查数据混用。

但是，对这一特性进行长期研究后你就会发现，群体预测通常会打败最佳的个体预测。例如，一份关于“蓝筹经济指标”的调查研究显示，在多年期内，由 70 名经济专家组成的智囊团中任何一名专家发表的预测都没有群体预测准确。沃尔弗斯还作过另外一项研究，该研究关注了美国国家足球联盟比赛的相关预测，发现由博彩市场做出的群体预测要比 99.5%的个体预测准确。这个规律同样适用于政治民调，任何一种将某项民调结果当作“圣杯”的模型都更容易遭遇尴尬的失败。群体预测能使错误率降低 15%~20%，这个幅度听起来似乎还不足以令人瞠目，但是在竞争激烈的市场中，群体智慧确实很难被打败。

所以，我告诉沃尔弗斯和罗斯柴尔德，尽管我并不接受他们表述的所有细节，但打算接受他们结论背后的原理，承认群体预测的优点。毕

竟 Intrade. com 网站的下注者在预测时不仅能够利用“538”网站的预测结果，也能使用任何他们认为与之相关的预测（比如我们竞争对手的预测，其中有些还是相当不错的）。当然，下注者会带着“有色眼镜”看待这些信息，于是他们便会陷入困境。但“538”网站的预测以及其他任何人所作的预测，都不应该因此备受责难。

见我竟然愿意做出如此大的让步，沃尔弗斯似乎有些失望。但在我本人看来，如果无法肯定自己能打败 Intrade. com 网站，那我为何不承认它更胜一筹，转而采纳它的预测结果呢？

“坦白说，我对你的反应感到十分惊讶，”沃尔弗斯对我说，“如果有能打败 Intrade. com 的方法，那就打败它，不然你所作的一切意义何在？”

是的，我做这一切的动机是什么呢？首先，我发现预测是件智力趣事，还能提高我博客的访问量。

其次，我虽从预测市场的理论中获益，但当时还不知道 Intrade. com 网站这样的政治博彩市场已经如此成熟，当然，竞争标准还是比较低的。Intrade. com 网站确实越来越受欢迎了，但和股票市场或拉斯韦加斯赌场相比，还是“小巫见大巫”。比如，在 2012 年 3 月美国共和党“超级星期二”选举开始的前几周，Intrade. com 网站的股票交易总额为 160 万美元，与之形成鲜明对比的是，纽约股票证券交易所每秒的交易额就达到了 800 万美元。“超级星期二”选举的最大赢家获得的利润约为 9 000 美元，还不够他维持日常生活，更不要指望靠此发家。同时，Intrade. com 网站还处于法律的“灰色地带”，大多数对美国政治下注的人都来自欧洲或其他国家。该网站还发生过一些操纵市场的事件，或存在明显不合理的赌博金额。另外，这些博彩市场并不是很擅长综合信息，因为在现实世界中并没有那么多有价值的信息可供收集，比如预测最高法院审判的结果，美国司法部门提供给公众的线索就十分模糊。

如果政治预测在美国完全合法化，而且其交易量能高出那么一两个数量级，那“538”网站及其他比较出色的政治预测网站是不是就能打败 Intrade. com 网站呢？我认为要做到这一点比较难。那现在，这些政治预测网站能打败 Intrade. com 网站吗？对于这个问题，我会给出一个有根据的推测，如果认真选择下赌注的内容，这些网站还是能够打败它的。

鉴于此，我想再次重申，当许多智者认为市场是可预测的，认为自己能够打败整个市场时，往往都遭遇了“滑铁卢”。

有效市场假说理论的缘起

1959 年，20 岁的大学生尤金·法玛受够了塔夫斯大学的罗曼斯语课程和有关伏尔泰的课程，于是他开始替一位教授打工，这名教授经营着一家股票市场预测服务机构。这份工作十分适合法玛。法玛天生就是一个强大的竞争者，他是家里第一个考进大学的人，虽然身高只有 1.77 米，却成了波士顿莫尔登天主教高中的明星运动员。工作中，法玛负责搜集股票以往的收益情况，从中寻找一些能为投资者提供帮助的常见数据模型，如果真能找出这些模型，那就表明股票市场具有高度的可预测性，投资者也能利用这些模型创造财富。但教授对法玛的策略总是充满疑虑，建议法玛先观察一下这些策略在现实世界中的实施效果，然后再着手投资。结果，法玛的策略总以失败告终。

这种经历让法玛备受打击，但也倍感兴奋。于是，法玛放弃了成为高中老师的计划，而是考取了芝加哥大学商学院。1965 年，法玛发表了自己的博士论文，这篇论文颇有棒球统计奇才比尔·詹姆斯在 20 世纪 80 年代的那种开拓型风范。法玛依靠一系列混合数据，略带讥讽地宣称，那些股票市场操作的传统经验完全是一派胡言。法玛研究了 1950~1960 年间的数十种股票的收益，发现某一年收益好的股票到了次年就无法击败同类竞争股票。事实就是，法玛先前的尝试证实了，他无法击败整个市场，别人同样也不行：

> 一个优秀的分析师的股票收益通常会比市场好。“通常”是这句话的关键词，因为在短时期内有人能优于市场，有人却略逊一筹。
>
> 不幸的是，根据这一个标准，本书作者也算不上优秀的分析师。但令人略感欣慰的是，其他市场测评机构似乎同样也不合格。
>
> 尤金·法玛，《股市价格的行为》

尽管这篇论文后来被引用了 4 000 多次，但起初它和芝加哥大学其他研究生发表的文章一样名不见经传。这篇论文奠定了有效市场假说理论的基础，该理论的核心论断认为，股市的运转在很大程度上是不可预测的。当然，在短时期内一些投资者表现得更好，这就好比某晚在拉斯韦加斯的轮盘赌桌上，某些赌徒能赢得盆满钵满。但是法玛认为，从长期来看，预测者仍无法击败市场。

过去的表现不代表未来的结果

很多时候，我们都没有觉察到有限样本存在的局限性，或者在评判某人的预测结果时误将运气视为技能。当然，把技能当作运气的情况时有发生，例如，在检验某位棒球手短期内的平均击球率时，其实他的棒

球技能尚佳，主要也是凭借技能得分，但噪声干扰太强时技能很容易被视为运气。

在股票市场，有关个体投资者表现的数据也充满了噪声，让人难以分辨他们的表现是好是坏。于是，“过去的表现不代表未来的结果”这句话出现在共同基金的宣传小册上是有道理的。

假设 2007 年你想投资一只共同基金，你主要关注的是美国股市的大盘股，如道·琼斯工业平均指数或标准普尔 500 指数所涵盖的股票。你咨询了亿创理财公司，该公司向你提供了几百只这类基金及其各方面信息，比如某只基金过去 5 年里的平均收益。当然，你最好投资 EVTMX（伊顿万斯分红基金 A）这类基金，该基金在 2002~2006 年以 10%的比例连续领先市场平均值。如果你愿冒风险，可选择 JSVAX（杰纳斯反向基金 T），该基金选择的股票不太受大家欢迎，但在同一时期内其收益每年以 9%的比例高于市场平均值。

其实，两者之间没有什么太大的差异。我对这些共同基金在 2002~2006 年的表现进行回顾，并将这些基金在 2007~2011 年的表现与前一时期做了对比，发现这些基金在不同时期的表现几乎毫无关联。EVTMX 在 2002~2006 年被评为最佳表现基金，但在随后 5 年里却表现平平。表现优异的 JSVAX 在接下来的几年却以 3%的比例连年低于市场平均值。正如法玛发现的那样，没有一只基金能一如既往地表现良好，甚至 5 年期限就是极限。其他研究也表明，共同基金每年的表现并非紧密相关，但也不是完全割裂开来的，所以最好的投资策略就是选择一只专业服务费用最低的基金，或是完全不理会理财公司而独自在股票市场中闯荡。

技术分析法并不能预测股市

然而，法玛对被他称为“技术分析者”的指责已经算口下留情了。“技术分析者”称，仅仅凭借过去的统计模式，而不用担心公司的盈亏，也不必管它是卖飞机还是卖汉堡，他们就能预测股市的走向（这正是法玛之前曾为之努力却遭遇失败的尝试，这类方法有一个更含蓄的叫法——技术分析。）

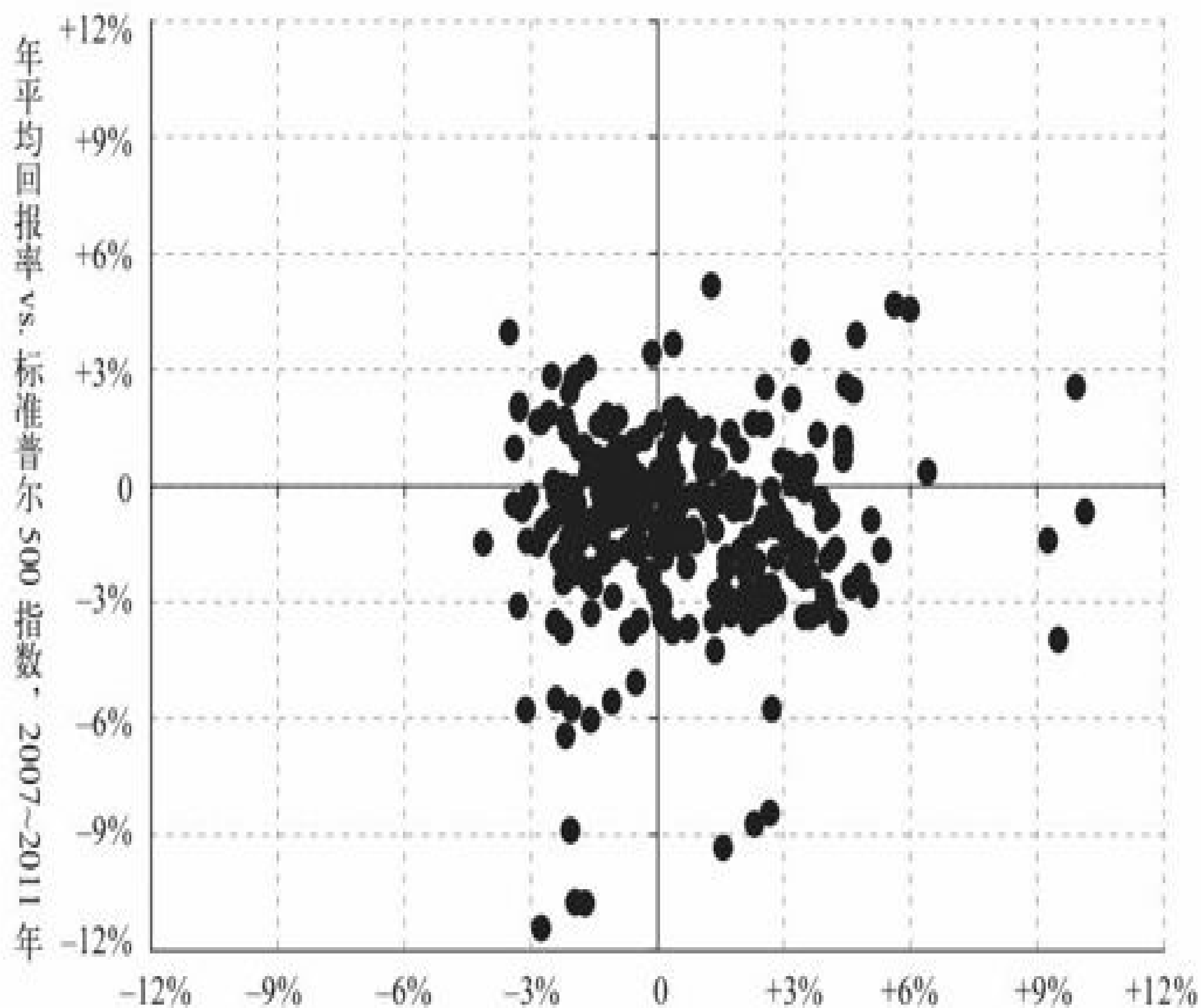
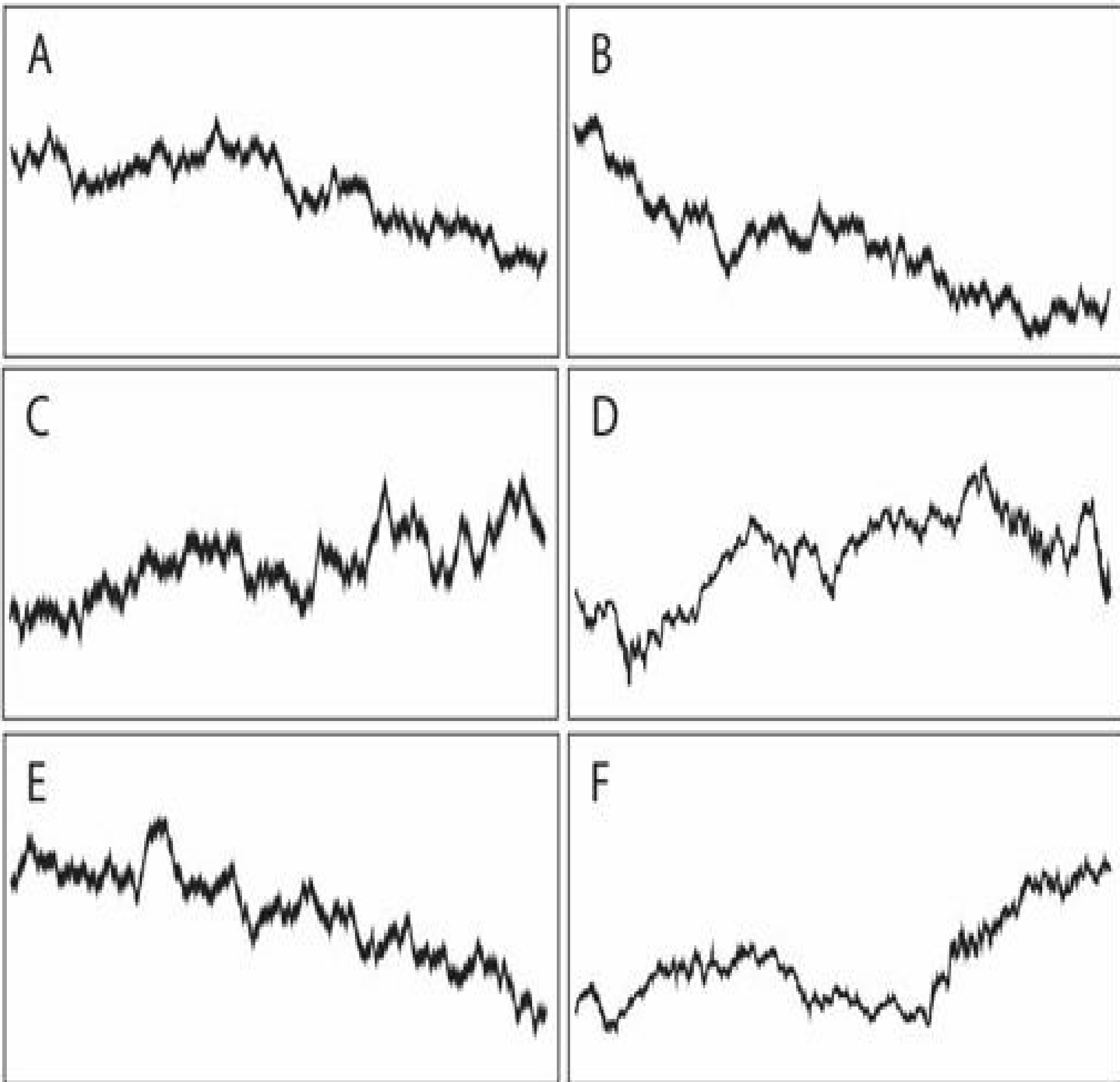


图 **11–3** 共同基金前后矛盾的表现

也许，我们应该对可怜的“技术分析者”心存同情，因为要从噪声中分辨信号实属不易。在图 11–4 中，我呈现了 6 幅股票价格走势图，其中 4 幅走势图是伪造的，是我用电脑随机生成的（或者，只是由一系列 1 和 0 这样的变量组成的）。另外两幅走势图则分别描述了 20 世纪 70 年代和 20 世纪 80 年代道·琼斯工业平均指数前 1 000 个交易日的真实走势。你能从 6 幅图中找出这两幅图来吗？显然这并非易事。投资者观察的股价走势也是如此，常常都将噪声误认为信号。



有效市场假说的 3 种形式

观察了上述类型的数据之后，法玛对有效市场假说作了修改，以便覆盖以下 3 种特殊形态，它们对市场不可预测性的表述越来越大胆。

第一种是弱式有效市场假说。该假说宣称，仅仅通过分析过去的统计模式是无法预测股价的。换言之，技术分析者的那些技术注定失败。

第二种是半强式有效市场假说。这种假说更进一步，认为基本分析法——观察公司的公开信息，包括公司财务状况、运营模式、宏观经济情况等——注定会失败，这种方法带来的回报通常无法打败市场（回报）。

第三种是强式有效市场假说。这一假说甚至宣称小道消息——内部

交易秘密——与市场价格密切相关，但无法带来超常的回报。这一版本的有效市场假说被视为该理论的逻辑极端，大多数假说支持者（包括法玛自己）基本上都不认同这一点。相反，有确切证据表明，内部人士都可以获得超常的收益。有一个令人不安的例子，即某些公司在游说美国国会议员时通常会将公司的内部信息透漏给他们，这些议员也可以通过立法影响这些公司的命运，他们的投资收益每年都会高出市场平均值5%~10%，如此高的收益率甚至让伯尼·麦道夫这样的野心家都有些脸红。

有关弱式和半强式有效市场假说的讨论，已经成为社会科学中最热门的话题。每年约有 900 篇关于有效市场假说的学术论文发表，现在的金融类期刊对这一话题的讨论热度与生物学界对进化论的讨论热度类似。

有效市场假说常被误以为华尔街为自己的越轨行为寻找的借口，因为这个假说似乎在说，无论华尔街这些家伙在做什么，至少他们的行为是理性的。但其实，有效市场假说的支持者很少会这样解读这一假说。该理论建立的初衷恰恰相反：股票市场无论是在深度还是广度上都是无法预测的。当某件事物确实无法预测时，无论是美发师还是年薪为 200 万美元的投资银行家都无法成为常胜将军。

然而，尽管自诩“威力无边”，但该理论同样带有限制条件。其中最重要的一个限制条件就是，理论中提到的回报是随风险调整的。假设你选择的投资策略每年需承担 10%的破产风险，如果未来 20 年你都坚持采用这一策略，那实属愚蠢之举，因为你的投资成本维持到这个时限的概率只有 12%。但是，如果你真有如此胆识，那就应该获得超额利润。上述 3 个版本的有效市场假说都留给投资者获得超过平均收益的空间，但相应的风险也会增加。正应了那句老话，收益越高，风险越大。

另一个重要的限制条件是，有效市场假说所提及的利润是指扣除交易成本的净利润。投资者每次交易股票都会产生交易成本。在大多数情况下，这些成本都比较小，约占每笔交易的 0.25%。但是，你交易得越多，交易成本累积得越多，对一个过度活跃的投资者来说，这个费用会是一笔很大的负担。这一点给了有效市场假说一定的缓冲空间。在无交易成本的地方，许多投资策略都能够产生薄利。但是在现实生活中，一个投资者需要赚足够多的钱才能支付这些额外开支，就如同扑克牌玩家得赢很多钱才足以支付一间游戏室的使用费。

追涨杀跌的股市投资策略真能获利吗？

有效市场假说的反对者有两种有效途径反驳这一假说。一种有效途

径是证实有这样一些投资者，他们的赢利总是超过股票市场的平均利润。另一种有效途径则更为直接，即表明股票收益是可预测的。

有一种简单的方法可以驳倒有效市场假说，那就是证明每日的股票价格波动是相互关联的。如果星期二那天股价上扬，是否意味着到了星期三股价也会走高？如果是这样，投资者可以在股价升高时买进股票，在股价下跌时抛出或空仓，这样一个简单的策略就能帮助投资者获利。当然，这还要取决于他的交易成本的金额。

假设 1966~1975 年的 10 年里——正好是法玛毕业论文发表后的 10 年——我们每天都观察道·琼斯工业平均指数的收盘价格，其间道·琼斯工业平均指数的变动方向在 58%的时间里都能保持一致——赢利日后接着赢利日或者亏损日后接着亏损日，在另外 42%的时间里股市走势发生变动。这一切似乎不是随机发生的，一项标准的统计测试宣称，这一现象纯属偶然的概率极低，大约只有 1/7 000 000 000 000 000。

但是，统计上的显著性并不等同于现实中的显著性。一个投资者并不能从这种趋势中获利，我们可以举例证明这一点。

假设一位投资者 10 年来持续观察这一模式——赢利跟着赢利或者亏损跟着亏损。1976 年 1 月 2 日上午，他决定将 10 000 美元投进一只紧随道·琼斯工业平均指数走势变动的指数型证券投资基金。这位投资者并不属于被动的投资者，他决定利用这一模式，采取“追涨杀跌”的投资策略。每当股票价格下跌时，这位投资者就会抛售所有的股票，因为该模式预测出次日股票的价格会继续下跌。观察到股市开始回暖时，他才会把钱再次投入股市。10 年来，他一直采取这种策略进行投资，直到 1985 年最后一个交易日，他十分肯定会获得大笔利润，于是便把所有的股票套现了。

这位投资者 10 年间究竟赚了多少钱呢？如果忽略分红、通货膨胀及交易成本等因素，凭借这个投资策略，这位投资者 1976 年投资的 10 000 美元经过 10 年会变成 25 000 美元。而另外有一位投资者采取了“买入并持有”这一简单的投资策略，10 年里从未中途买入或卖出，最初投入的 10 000 美元最后变成 18 000 美元。两者相比，看来“追涨杀跌”的投资策略似乎奏效了。第一位投资者利用以往股价走势中一个简单的统计学关系，采取“追涨杀跌”策略的而获利，这一点似乎能够推翻有效市场假说。

但有一个问题：我们忽略了这位投资者的交易成本。这会使结果大不相同。假设第一位投资者仍旧采取“追涨杀跌”的策略，但每一次买入或卖出时，他都得拿出交易额的 0.25%作为佣金给经纪人。由于这项策略共需要交易上百次，这些看似小额的佣金却会一点点地把他的利

润“榨干”。事实上，如果扣除该策略 10 年产生的交易成本总额，10 000 美元的投资最终就只剩下 1 100 美元了，不但没有获得任何利润，就连最初投入的那笔钱也几乎损失殆尽。在这个案例中，股市收益确实具有一定的可预测性，但还不足以凭此获利，所以有效市场假说仍然成立。

还有一个问题，即该模式会自我消退。在 21 世纪初的 54%的时间里，股市走势都发生了变动，而没有遵循“赢利跟着赢利或者亏损跟着亏损”这一模式，这和前一个 10 年的情况正好相反。在这样的股市环境中，如果某位投资者从 2000 年 1 月起采取“追涨杀跌”的投资策略，10 年后，在没有扣除交易成本的情况下，原来 10 000 美元的投资就只能剩下 4 000 美元了。若是再扣除交易成本，就只剩下可怜的 141 美元了，几乎 99%的本金都亏掉了。

这个案例就是告诉大家，请不要尝试这种投资策略。这种策略与“石头剪子布”游戏[1]十分相似，风险都极高，它所产生的交易成本将会吞噬你的利润，还会侵蚀你的本金。正如法玛与他的教授觉察到的一样，那些看似太过准确而显得不真实的股市投资策略通常是不真实的。如同预测地震发生频率的历史模型一样，股市似乎就像一座炼狱——它的数据既不完全随机，也不完全不可预测。然而，股市的情况更加棘手，因为它的数据不是用来描绘某种自然现象，而是人类的集体行为。如果你在探寻一种模型，尤其是那些看似显而易见的模型，那么你找到这种模型的概率和其他投资者是相同的，模型中的信号会抵消噪声，甚至还会出现自行消退的状况。

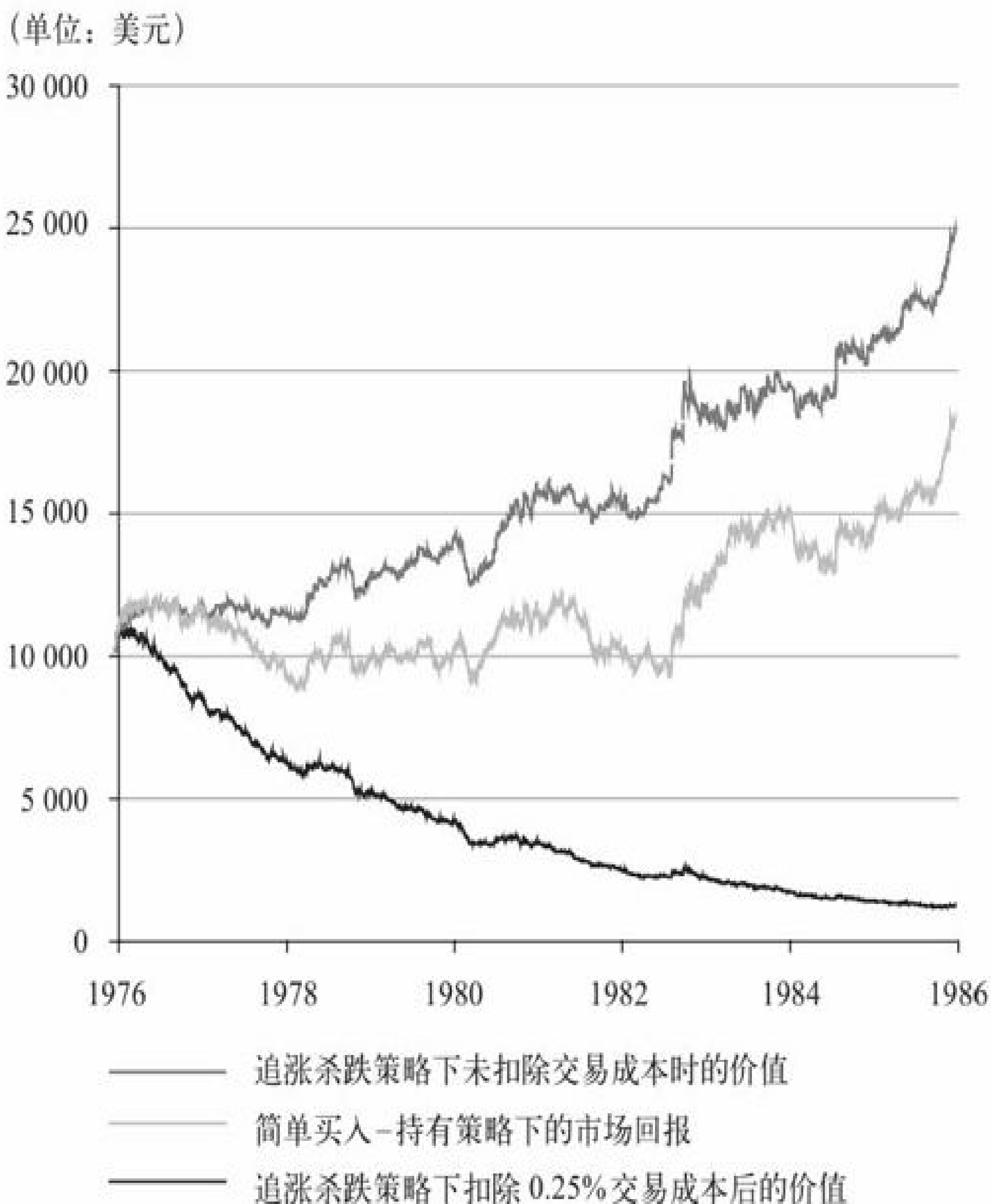


图 **11-5** 扣除交易成本和未扣除交易成本的股票价值曲线图

当有效市场假说遇到非理性繁荣

另一个对有效市场假说的有力挑战来自持续上扬的股价，比如 20 世纪 90 年代和 21 世纪早期科技股价格的飞涨。1998 年年末到 2000 年年

初，纳斯达克综合指数比之前高 2 倍多，但是此后两年，所有（以及另外一些）收益便全部散尽了。



图 **11-6 1990-2004** 年纳斯达克综合指数

一些在纳斯达克证券交易所上市的股票价格，看似明显不合理。在网络公司日益繁荣的时候，科技公司的市场价值约占美国所有股票市值的 35%，这意味着这些科技公司的产值很快就会占到私有部门总利润的 1/3。有趣的是，科技本身的发展确实大大出乎众人意料。如果在 2000 年向一位投资者展示苹果平板电脑，并告诉她在未来 10 年内，当她飞行在密苏里州 10 500 米高的上空时可以通过苹果平板电脑浏览网页，可以用 Skype[2]和远在中国香港的家人通话，她听后会有何反应？这位投资者肯定会说苹果公司的股价将会“钱”途无限吧。

然而 10 年之后，到 2010 年，科技公司的产值贡献只占所有经济活动总量的 7%。类似苹果公司这样的企业多出一家，就会有数十家 Pet.com 这类公司破产。但是，投资者似乎觉得每一家科技公司都会成为赢家，认为它们之间不存在相互竞争，于是他们会产生一种不切实际的想法：整个科技行业都将获利。

但是，一些有效市场假说的支持者对泡沫这一概念仍旧带有抵触情

绪。有一次我和法玛闲谈，聊别的话题时气氛还很融洽，但一谈到“泡沫”这个词法玛就避而不谈。法玛语气强硬地对我说：“市场泡沫这个术语已经失去了原本的意义，市场泡沫的结果应该是可以预测到的。如果你都无法辨别是否身处泡沫中，那它就不是泡沫。”

泡沫理论要怎样才能驳倒有效市场假说呢？这要求我们能够做到对市场泡沫的实时预测。如果能做到，一些投资者就可以在泡沫出现时将其识别出来，并利用这种泡沫获取利润。

事后识别市场泡沫当然比较容易，但坦白说，提前识别市场泡沫也没有多困难，许多经济学家在房地产泡沫发生之初就已经看出苗头来了。通过简单的观察，若在一段时期内股票市场的增长率要比其历史平均水平高很多，这就暗示了股市泡沫的存在。标准普尔 500 指数的总价比其长期以来的平均值高出两倍，5 年里这样的状况共发生了 8 次，其中 5 次都伴随着严重的金融危机，比如经济大萧条、互联网泡沫以及 1987 年的“黑色星期一”。

另外一种更准确但十分复杂的泡沫检测方法是由耶鲁大学的经济学家罗伯特·希勒提出的，有关希勒对房地产泡沫的预见性研究，在本书第一章中已讨论过。希勒最知名的著作就是《非理性繁荣》。该书出版时正值纳斯达克指数达到互联网泡沫的顶峰时期，它的出版是对其他书籍的纠正，如《道·琼斯工业平均指数 36 000 点》、《道·琼斯工业平均指数 40 000 点》、《道·琼斯工业平均指数 100 000 点》等，这些书向投资者承诺股价会持续走高，却没有警告他们当时股价已脱离了定价的原则，虚高的情况很严重。

从理论上讲，股票价值终究是对一家公司未来收益和分红的预测值。尽管未来收益难以预测，但可以参考这家公司近几年的收益（希勒的公式参考的是过去 10 年的收益），并将其与该公司的股票价值进行比较。用公式表达就是 P/E，即股价收益比，也称市盈率，若计算得出某家公司长期以来的股价收益比为 15，就意味着该公司股票的价值是其年利润的 15 倍。

个股中会有例外情况，有时这种例外情况还是合情合理的。新兴产业的某家公司（如脸谱网）未来的收益比之前多，这就是合理的。因此，这类公司的市盈率自然会高于没落产业的某家公司（比如曾经的影片租赁巨头百视达影视公司）。然而，希勒所观察的对象是标准普尔 500 指数涵盖的所有上市公司的平均市盈率。从理论上讲，纵观企业的平均市盈率，新兴产业中高市盈率的公司与没落产业中低市盈率的公司数量相当，所以整个股票市场的市盈率可以长期保持基本稳定。

但是，希勒发现事实并非如此。在不同时期，标准普尔 500 指数中

所有上市公司的市盈率变化幅度从 5（1921 年）到 44（2000 年希勒的书出版时）不等。希勒发现这些反常数据似乎具有预测价值，能为投资者所用。当市盈率为 10 时，这就意味着股价收益比较低，且以往都有约 9%的实际收益，这就意味着 10 000 美元的投资在 10 年后总值能达到 22 000 美元。然而，当市盈率为 25 时，根据历史推断，10 000 美元的投资在 10 年后却只能达到 12 000 美元。当市盈率很高，超过了 30 的时候——1929 年或 2000 年时就达到了 30——预期收益将会变成负数。

然而，除非你有十足的耐心，否则这一模式很难带来收益。只有从长期来看这些模式才有意义，不要期待它们在一个月后或是一年后就能告诉你市场价值有多少，因为那几乎是不可能做到的，甚至连提前几年做出预测都不可能，因为其预测力仍十分有限。比如，1996 年 12 月，艾伦·格林斯潘第一次使用“非理性繁荣”一词来描述技术股的状况，那时标准普尔 500 指数的市盈率已达到 28，这一数字离 1929 年的“黑色星期二”及“大萧条”时期 33 的市盈率已经不远了。按理来说，此时不应该再买入股票，但当时纳斯达克综合指数总值仍居高不下，科技股票泡沫的顶峰还要 3 年才会到来。若有位投资者有着精准的预测力，在格林斯潘发表那篇演讲的当天买入纳斯达克综合指数，而且抛售的时机把握恰当，就能赚回约为成本 4 倍的利润。然而，在一般情况下，确实只有达到 10 年或 20 年的期限，市盈率才能帮助投资者做出可靠的预测。

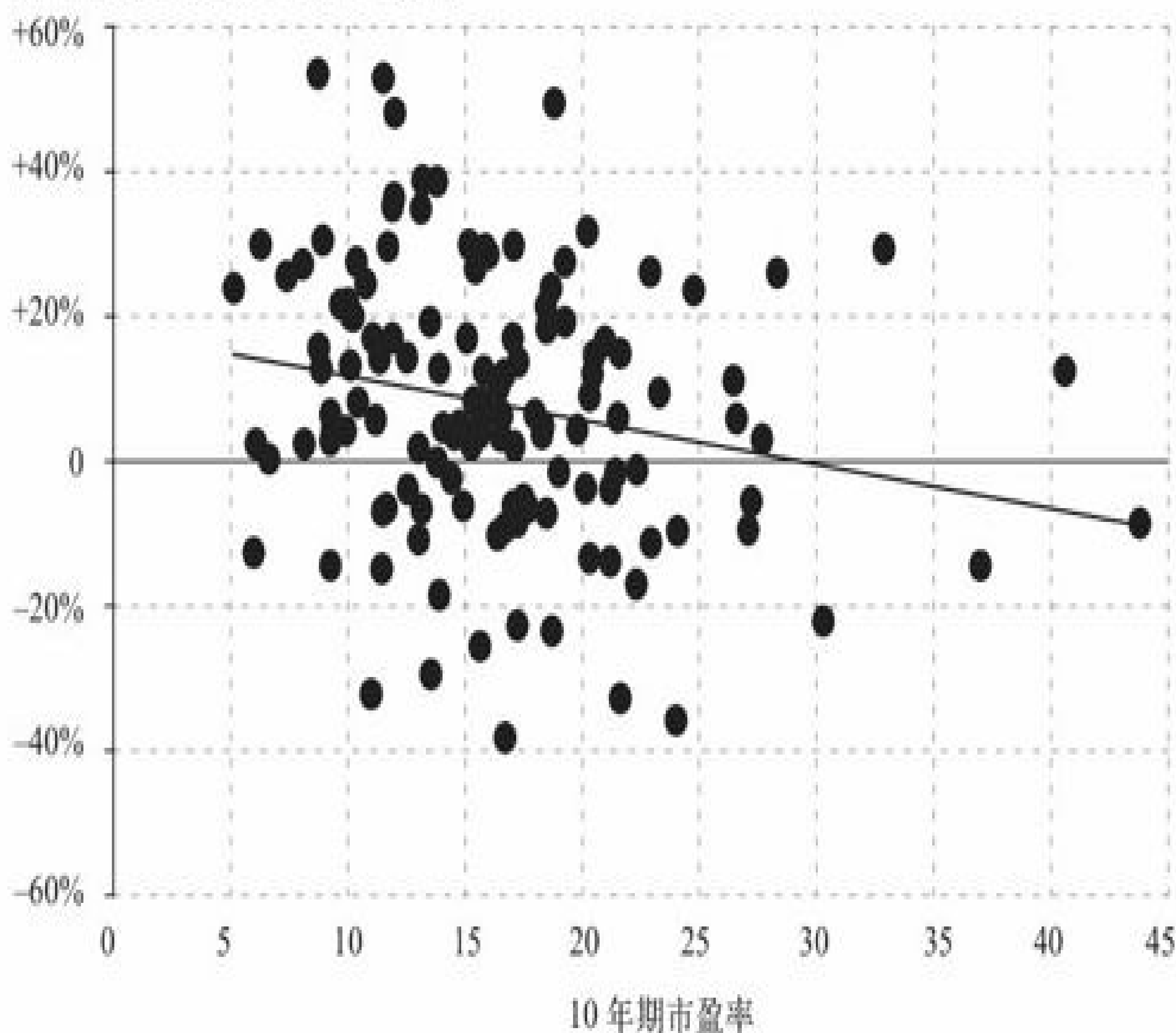
股票市场中几乎没有真正确定的事情[3]，这一模式甚至能够体现出信号和噪声的某种组合。然而，希勒的发现不仅有强大的理论支持，也有确切的实证支持，因为他是结合股票市场评估的基本原则来关注市盈率的，这就使得整个求取比例的过程更趋向真实。

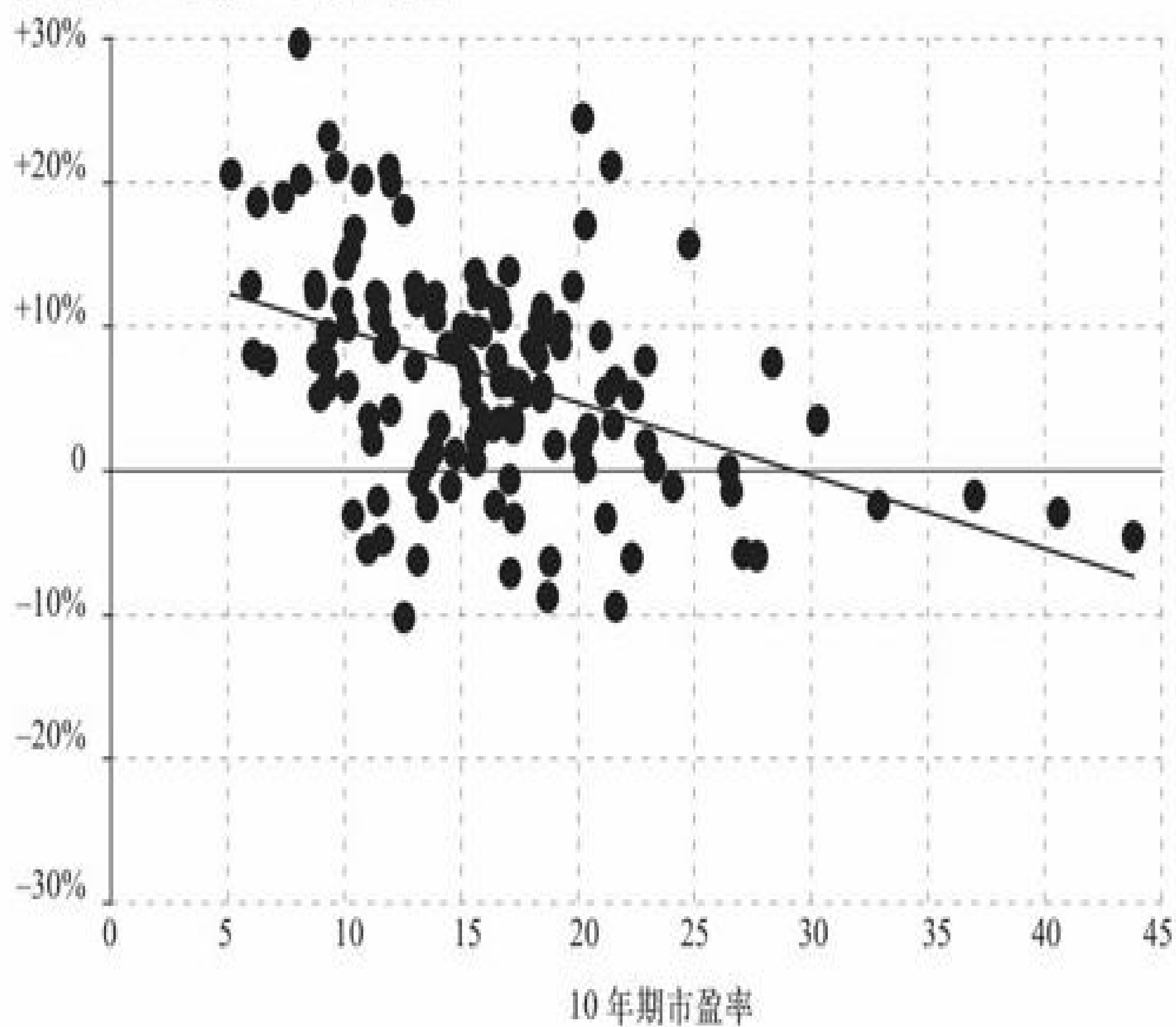
如果股价在短期内如此不可预测，那长期来看为什么又变得可预测了呢？答案在于股票经纪人在面对竞争压力时采取的行动，这些压力来自竞争的公司，也来自其客户和老板。

有效市场假说的大多数理论诉求就是认为（如同贝叶斯所认为的）股市中的错误应该自行修正。假设你观察到美国大型博彩公司——米高梅国际度假村集团（MGM）——的股价一到周五就会上升 10%，也许是因为股票经纪人潜意识中都渴望在周末到以赌闻名的大西洋城挥霍一番。在一个特别的星期五，MGM 开盘价为每股 100 美元，而你期望在交易日结束时这个价格能升到每股 110 美元，那么应该怎么做呢？当然先得买入股票，并期待这只股票迅速赢利。但是，当你买入这只股票时股价开始上涨。一笔大交易可能足以让股价从每股 100 美元升至每股

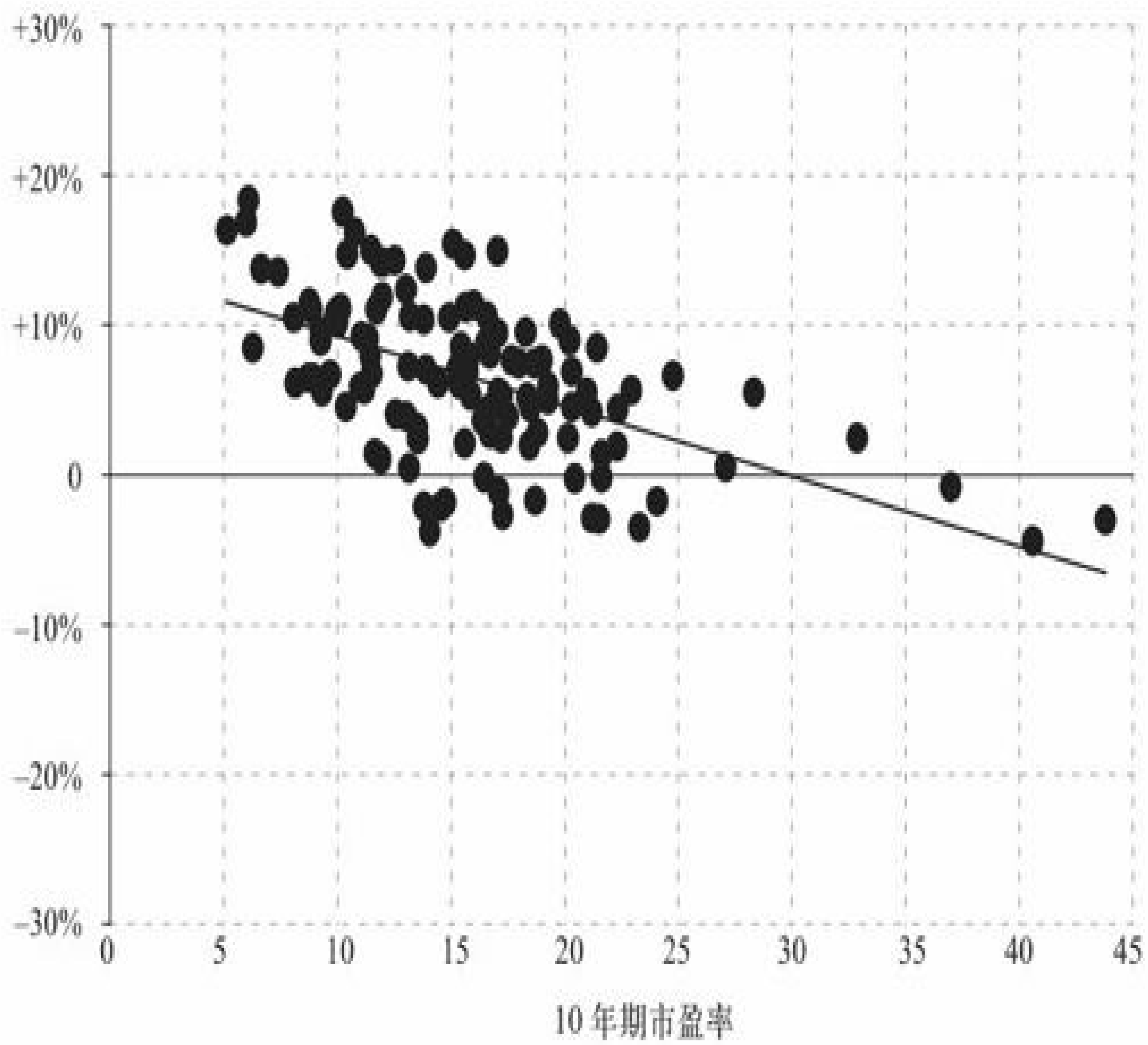
102 美元。但是，还是会有利润空间，于是你决定继续买入，此时股价升到每股 104 美元了。就这样持续不断，直到股价涨到每股 110 美元。回顾发生的一切：当你觉察到股价异常时，你已经在尽量消除这种异常了。

在现实世界中，这些模式绝对不会这样明显。股市中有上百万名股票交易者，包括数百名专注于博彩行业的分析师，怎么可能只有你一个人注意到每周五晚上 MGM 的股价会上涨 10%呢？然而，你通常只是在争夺一些剩余的利润，却很少反思：这样的统计模式究竟有没有意义？未来是否会持续存在？所带来的利润是否够支付交易费用？其他参与竞争的投资者是否也在利用这一模式？然而，所有竞争都意味着股市应该尽快适应这些错误的定价，而那些小错误是不值得担心的。至少，这一理论是这么认为的。





标准普尔 500 指数 5 年期平均收益
+30%
+20%
+10%
0
-10%
-20%
-30%
0
5
10
15
20
25
30
35
40
45
10 年期市盈率



标准普尔 500 指数 10 年期平均收益
+30%
+20%
+10%
0
–10%
–20%
–30%
0
5
10
15
20
25
30
35
40
45
10 年期市盈率

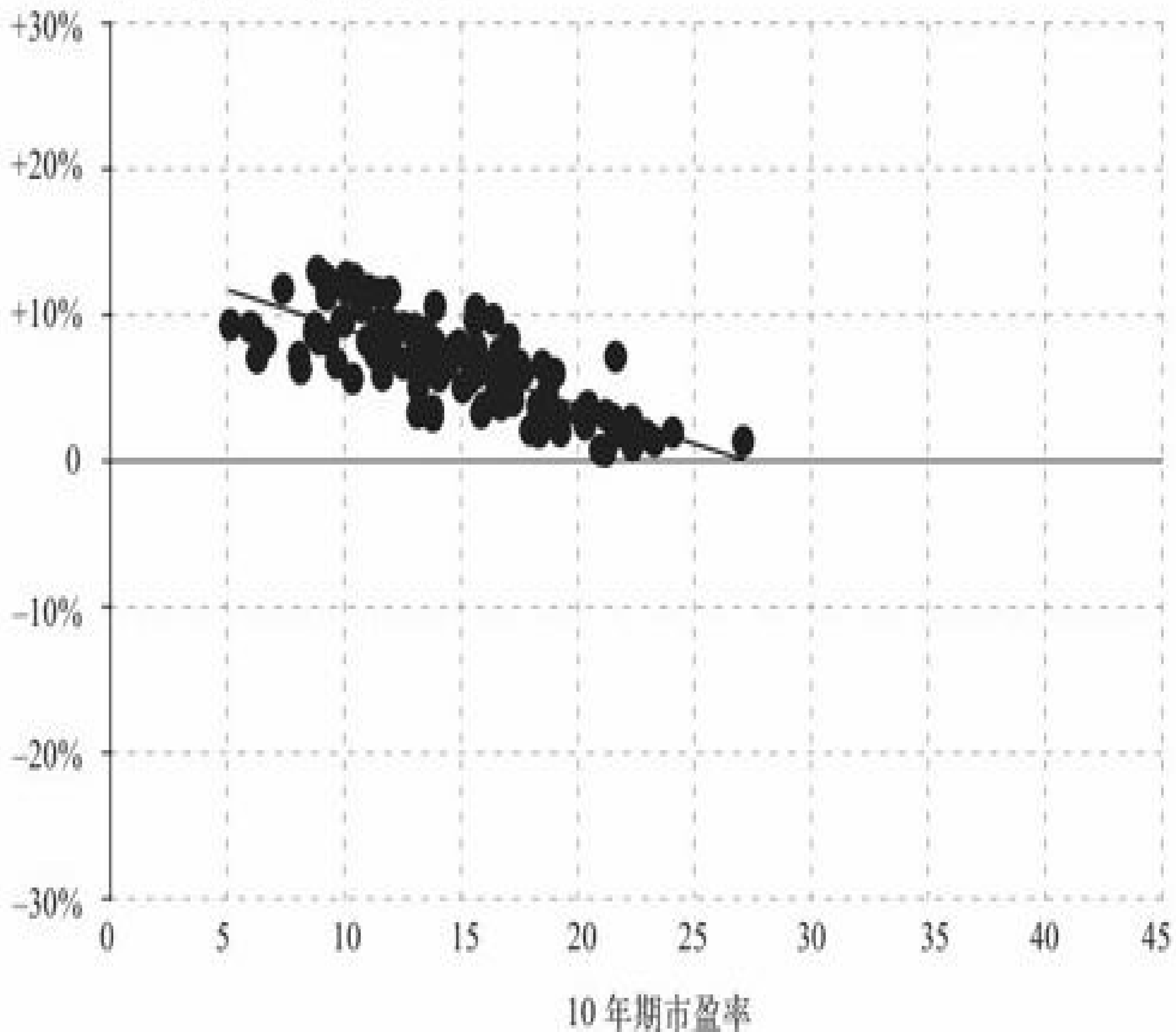


图 **11-7** 市盈率和股票收益

大多数股票交易者，尤其是那些最为活跃的交易者，都只着眼于短期目标。他们不会放过任何一个获取利润的机会，哪怕是要求他们提前一天、一个月乃至一年就开始思索获利的方法都无妨，但是他们并不在乎他们的行为产生的影响。也许股票市场存在其他一些可预测性，但在股票经纪人眼里，这不是他们的工作职责。

羊群效应催生股市泡沫

亨利·布洛杰特第一次受到全美国人的关注是在 1998 年。此前，布洛杰特作为一名自由记者兼英语教师，曾游历日本多年，回到美国后，在 CIBC 奥本海默公司担任互联网股票分析师。随着互联网股票泡沫越来越受到关注，布洛杰特的分析也开始为人们所关注。1988 年 12 月，

布洛杰特发布了一个大胆的预测：亚马逊网站的股价会在一年内从每股243 美元升至每股 400 美元。事实上，亚马逊的股价在此后不到两周的时间内就超过了 400 美元。

在当时那样一个疯狂的时期，布洛杰特的预测也许是一个自我实现的预言：在布洛杰特向大家推荐亚马逊网站的股票后，亚马逊网站的股价几个小时内就暴涨 25%。这个大胆的预测为布洛杰特赢得了声誉，并得到了美林公司分析师的工作机会，年薪高达几百万美元。布洛杰特颇具天赋，寥寥数语就能概括出股市的时代特点。布洛杰特是这么评说 1998 年的互联网泡沫的，“股票投资者购买的其实是他们对未来的憧憬”。如此能言善辩又声名在外，难怪布洛杰特会引得各类媒体竞相报道。

布洛杰特对亚马逊网站股价的预测，在今天来看都不失精准：1988 年他所推荐的亚马逊股票价格为每股 243 美元，到 2011 年时，若用相同的标准衡量，这只股票的价格会升到每股 1 300 美元。布洛杰特力劝投资者购买股票时应该看重公司的实际价值，更应关注一个行业的“领头羊”，比如，互联网行业的领头羊亚马逊网站、雅虎网站、易趣网站等，而不必太关注那些小型公司，用布洛杰特的话来说，这些小型公司“或被吞并，或走向破产，或逐渐消失在市场中”。在私有通信领域，那些营销策略有问题的小型公司，如 Lifeminder.com 网站有限公司、24/7 媒体公司、信息空间公司等并不在布洛杰特的推荐名单中，结果也证明这些公司确实毫无价值，市值缩水了 95%甚至更多。

可是有一个问题——一个大问题——出现了，不管布洛杰特私底下如何批评这些公司，但在公开场合，他还是在举荐 Lifeminder.com 网站这类公司，保留它们的买入评级，并在电视上为它们辩护。另外，如此表里不一的表现似乎有利于那些与美林公司有金融业务往来的公司。不久后，美国证券交易管理委员指控布洛杰特涉嫌诈骗，布洛杰特就案件的细节进行辩护，但最终还是被处罚金 400 万美元，终身不准再从事证券交易业务。

布洛杰特知道，现在无论他说什么，华尔街的人们都会心存疑虑；他曾为《石板》杂志写过一篇关于玛莎·斯图尔特审判的小文章，还附有一篇 1 021 字的披露声明。现在，布洛杰特有时间来细细品读经济学家法玛和希勒的著作了，还可以将书中内容和自己作为华尔街知情人的亲身经历进行对比。布洛杰特还开始了自己的记者生涯，担任知名博客帝国“商业内幕”网站的 CEO。所有这些经历，让布洛杰特在审视分析师和股票经纪人时尽管心灰意冷，但观点还算成熟。

布洛杰特对我说：“与很多投资经理人交谈后我会发现，他们可能会

考虑下周、下个月或是下个季度的情况。其实这份工作没有时间范围，关键在于和你的竞争对手相比，你自己当下的表现如何。确实，你只需要保证 90 天内不出问题即可，可一旦这 90 天内出现任何错误，客户就会辞退你，舆论媒体会挖苦你，从此以后你的表现会一塌糊涂。一旦这样，基本原则也救不了你。”

试想一下，若有位股票经纪人看过希勒的书，并接受了书中提到的基本前提——高市盈率是市值被高估的信号，结果会怎样呢？然而，经纪人只关心股票未来 90 天的表现。股市的市盈率高达 30，即股价估值已经是正常值的两倍，但在 90 天内股市崩盘的可能性只有大约 4%，这种情况在历史上也确实出现过。

如果这位股票经纪人的客户超有耐心，想要提前一整年预测股市的状况，他就会发现股市崩盘的概率会升至约 19%（图 11–8）。这和俄罗斯轮盘赌赌输一局的概率是一样的。这位股票经纪人知道，在引火烧身前这样的赌局只能玩几次。但现实是，他还有别的选择吗？

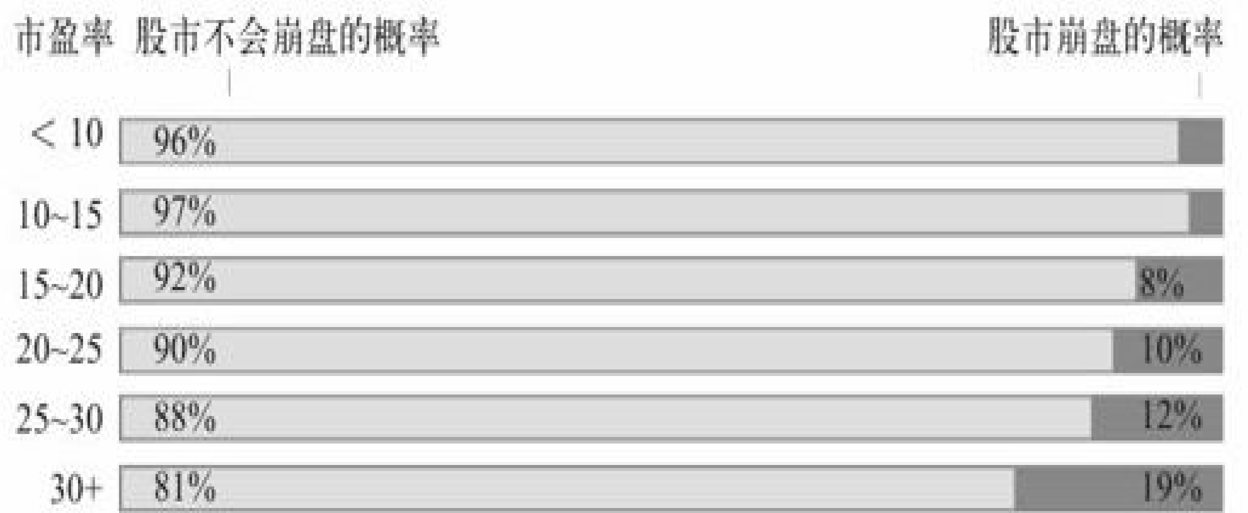


图 **11-8** 股票市场一年内崩盘的历史概率

布洛杰特只能做出决定：买入还是卖出。而股市可能会崩盘，也可能不会。于是，这个案例会有 4 种版本。首先是两个赌赢的版本：

- ·这位股票经纪人买入股票，并且股市上涨。该版本中一切正常。能从股市中获利，大家皆大欢喜。这位股票经纪人也得到了 6 位数的佣金，他用这笔钱买了一辆新款的雷克萨斯汽车。

- ·这位股票经纪人将股票卖出，而且股市确实崩盘了。如果他预计会发生崩盘，实际也确实发生了，再加上大多数人都赌输了，

此时他看起来简直就是个股市天才。接下来会有一个更好的工作机会等着他，比如，成为某个对冲基金的合伙人。然而，在股市崩盘、资金紧缩后，市场对天才的需求也是有限的。所以，很有可能随着媒体曝光率的增加，他获得的不再是上述各类机会，而是来自《华尔街日报》的约稿，约他写一篇受欢迎的评论，或是一个写书的邀约，又或是几个参会邀请，诸如此类。

你的个性会在很大程度上决定你的选择。选择第一种结局的人喜欢享受华尔街的生活格调，喜欢融入群体；第二种结局更适合那些喜欢打破常规的人。正如迈克尔·刘易斯在《大空头》[4]一书中所刻画的投资者一样，许多成功的投资者在 21 世纪初都敢于把宝押在抵押证券和其他泡沫投资上，但在一定程度上他们都不太懂人情世故，从上述第二种人来看，刘易斯这样的描绘并非巧合。

但是，设想一下如果投资者下错注了，情况又会怎样？他们面临的选择会简单得多。

·这位股票经纪人买入股票，但股市却崩盘了。这种情况真是不妙：他让公司损失了一大笔钱，自己也拿不到大笔佣金，不能买新款雷克萨斯汽车了。但他并不是“孤军奋战”，他的同事大多数都犯了同样的错误。回顾历史，经历了最近 3 次大危机的洗劫后，华尔街上那些证券公司的雇用率降低了 20%。这就意味着，这位经纪人还有 80%的机会能保住工作，结局尚可；等到下一次牛市，或许就可以把雷克萨斯汽车开回家了。

·这位股票经纪人将股票抛出，但是股价却涨了。这个情况对他来说不啻于一场灾难。这不仅使得同行们的操盘表现大大超过了他，而当初他竟还伸长脖子大呼同行们是傻子。结果是，他很有可能会被辞退。而且，以后也无法将自己推销出去，他再次受雇的概率微乎其微，可能获得的佣金也会骤减。

如果我是这位股票经纪人，当股市崩盘的概率为 20%的时候，我绝不会将股票抛出。甚至当这个概率达到 50%的时候，我也不会抛出股票。等到崩盘危情迫在眉睫时，我都还愿意铤而走险，但期望所有人和我同舟共济，生死与共。

的确，大型证券公司都倾向于避免“鹤立鸡群”，不到万不得已，不会降低某只股票的买入评级。2001 年 10 月，17 名分析师中有 15 名依然将安然公司的股票买入评级定为“买入”或是“强力买入”，尽管那时安

然公司的市值已经缩水了 50%，并曝出了财务丑闻。即使这些证券公司知道某家公司大势已去，但最有利于他们的选择仍旧是尽可能地维持这只“将死的骆驼”的生命。“我们以为这是第八局，没想到已经到了第九局了。”对冲基金经理史丹利·德鲁肯米勒于 2000 年 4 月对《纽约时报》如是说，当时他的量子基金的市值短短数月内就缩水了 22%。德鲁肯米勒已经知道科技股被高估了，股价必将下降，只是他没料到这场危机会来得这么快。

如今的股市中，大多数股票经纪人都在用他人的钱进行交易（拿德鲁肯米勒来说，他炒股的钱基本都是乔治·索罗斯的）。20 世纪 90 年代到 21 世纪初被看作日内交易者的时代。但是机构投资者，如共同基金、对冲基金及养老金，它们持有的股票资产增长更快（图 11–9）。20 世纪 60 年代法玛写论文的草稿时，机构投资者持有的股票只占市场的 15%，而到 2007 年，这一比例已经升至 68%。

这些统计数据代表了有效市场假说的潜在复杂性：若投入股市中的钱是别人的而不是你自己的，你的动机就不一样了。事实上，如果实际情况允许股票经纪人和其他同行站在同一个战壕里，而且可以降低他们被辞退的概率，那么这些股票经纪人即使给公司和投资者造成了损失，也在情理之中。有确凿的理论证据和实践经验可以证明，在共同基金的个体投资者和机构投资者中，存在羊群效应。布洛杰特告诉我：“为什么会形成市场泡沫，答案就是人人都期望股市能持续上涨以实现自己的利益。”

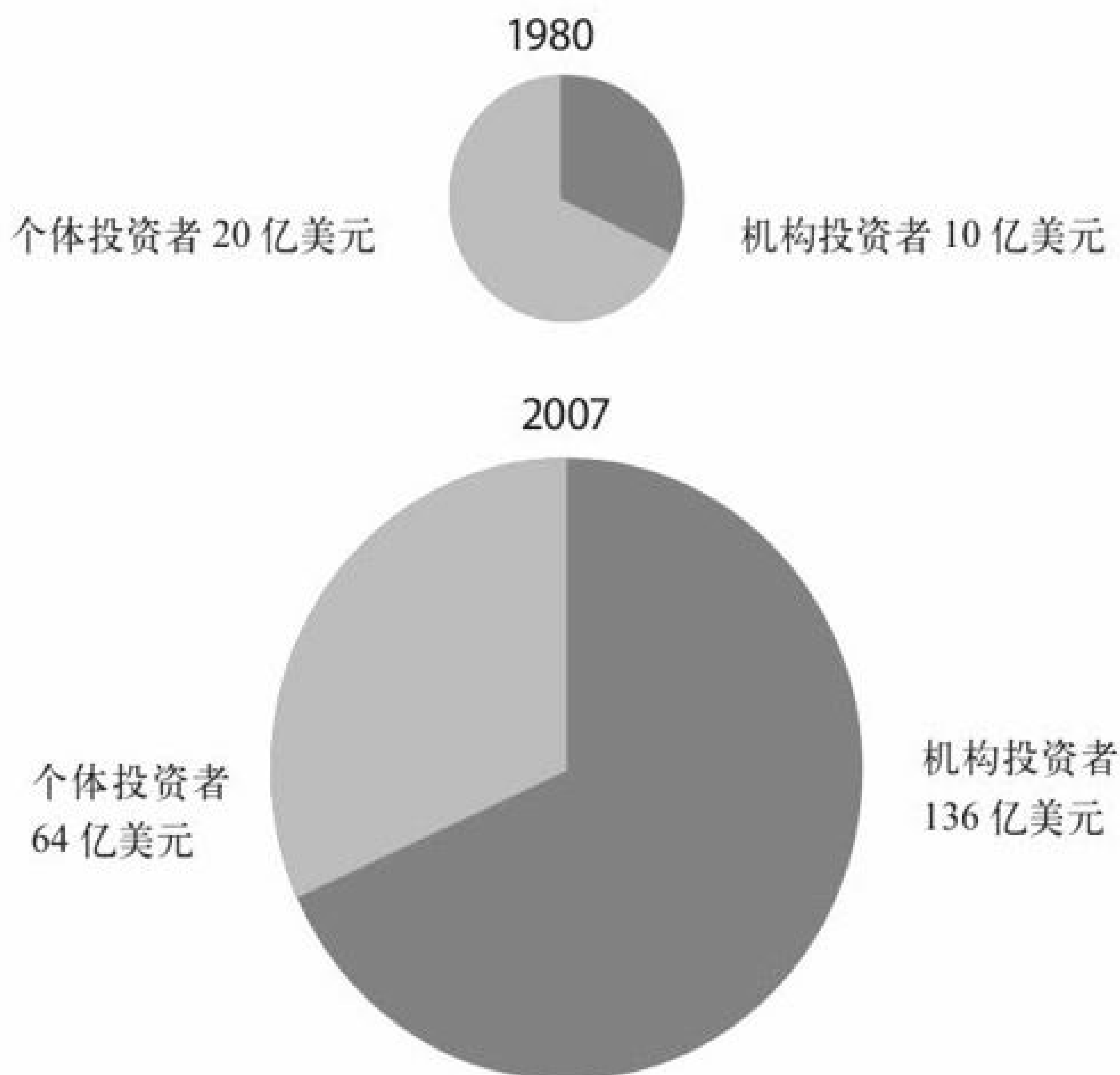


图 **11-9** 美国个体投资者和机构投资者的持股比（考虑通胀因素）

到目前为止，我所描述的一切现象都源自股市个体参与者的理性行为。投资者对待自己的投资都是高度理性的，但这不意味着他们的态度就一定能够帮助自己实现利润最大化。经济学中有这么一个观点，即使市场中的参与者的表现都是非理性的，市场作为一个整体也可以表现得十分理性。但是，市场中的非理性现象则可能源自个体依据自身动机做出的理性反应。只要人们评判大多数股票交易者的标准仍是他们的短期表现，股价严重偏离其长期市值所造成的市场泡沫就会时有发生，甚至是不可避免的。

羊群效应同样深受心理因素的制约。当遇到人生的重大抉择时，多数时候我们都会征询家人、邻居、同事和朋友的意见，有时甚至会问问

自己的竞争对手是否愿意给点意见。

假设我认为纳达尔赢得温网冠军的概率为 30%，但是我遇到的所有网球迷都认为纳达尔有 50%的概率赢得冠军，那么，我必须是一个非常自信的人才会坚守自己的观点。除非我有独家信息，例如他们都无法获得的信息，或者我确信自己比他们更用心地研究过这场比赛，否则在这种情况下，我这次有违常规的预测一定会失败。有一条启发法可谓百试不爽——“若无良计可施，那就随大流吧”。

然而，有时我们会过于信任邻居，就如美国 20 世纪 80 年代“请说不”的那则广告中说的一样，我们之所以这么做，是因为别人都在这么做。群体智慧的观点认为我们的错误会相互抵消，可实际却恰恰相反，它们开始彼此加强，逐渐失控。“瞎子引瞎子，二人掉深渊”，这种现象很少发生，可一旦发生便是灾难。

有时我们也许还会推测，那些最自信的邻居一定预测得最准确，于是便追随其后——不管自信的邻居是否真的知道自己在做什么。2008 年，不知出于什么原因，Intrade. com 网站一名不安分的股票经纪人无缘无故大量买入“麦卡恩概念股”，同时大量卖出“奥巴马概念股”。这就造成了股价异常，最终这些异常会被修正，但是每次都得花费 4~6 个小时股价才会最终回到原来的价位。许多投资者都相信这位股票经纪人一定掌握了一些独家信息，也许他了解一些即将被披露的丑闻？于是，众人开始纷纷效仿。

这就是从众效应（也称羊群效应）。有证据表明，这一现象在股市中越来越普遍。不同股票和不同资产之间的价格变动关系变得越来越清晰，这表明大家的投资范围广，各种投资品种都有所涉及，但都采取同一个投资策略。这也是信息爆炸时代的另一个风险：共享信息如此丰富，独立性反而被削弱了。

在股市中，有时股价是由那些最差劲儿的投资者引导的，只是因为他们的交易量最大。

过度自信的投资者逃不开“赢家的诅咒”

在经济学课上有这么一个常见的实验，通常教授想额外赚点午餐费时就会让学生做这个实验。他会在班内举行一场拍卖会，让学生竞拍一罐硬币。出价最高的学生将赢得那罐硬币（如果不想要零钱，他也可以得到等价的纸币），还要向老师支付一笔拍卖服务费。几乎每一次中标的学生都会发现，这笔交易不值那么多钱。有些学生报价过低，有些报价正好等于那罐硬币的金额，但是，每次中标的都是估价过高的那位同学，而最差的预测者往往会获得“奖励”。这就是赢家的诅咒。

股市也有同样的特性。有时，某个股票投资者十分乐意买入某只股票，因为他对这家公司有着自己独到的看法。但是，大多数投资者都处于平均水平，所用的模式都一样，采用的数据也相同。如果有些投资者认为某一只股票的价格过低，但其他投资者却不这么认为，那么可以肯定地说，大多数时候都是因为这些投资者对自己的预测能力过于自信，错误地将噪声当成信号了。

我们有理由怀疑投资者存在多种认知偏见，而过分自信就是其中危害最大的偏见。可以说，行为经济学的核心成果就是发现大多数人在预测时都表现得过分自信。股市中的预测也不例外，美国杜克大学一份关于公司首席财务官的调查发现，人们期待这些投资者能较为老练成熟，实际上却高估了他们对标准普尔 500 指数的预测能力。他们竟对股价的剧烈波动感到意外，完全忽视了长久以来股市在短期内都相当不稳定这一个不争的事实。

加利福尼亚大学伯克利分校的经济学家特伦斯·奥丁构建了一个模型，模型中的投资者都只有一个缺点：他们在评估信息的价值时都过于自信了。而在其他方面，他们都表现得十分理性。奥丁发现仅过度自信这一个缺点就足以扰乱理性市场。如果市场中充斥着一群过度自信的投资者，交易量便会激增，股价浮动幅度加大，每日股价间的关系反常，活跃投资者的收益也会低于平均利润，而所有这些在现实世界中都会发生。

股市泡沫需要很长时间才能被挤出

然而，如果发生股市泡沫的趋势明显，有效市场假说表明，某些投资者就应该能干预泡沫形成，并通过卖空股票赚取大笔利润。最终这个推论将会成立：所有泡沫都会被击碎。然而，这些泡沫事实上需要很长时间才能被挤出。

对付被高估的股票，最好的办法就是卖空，即高抛低补，具体来说是指股票投资者从交易商手中借入该股票抛出，在实际交割时以此刻的价格结清。若交割时价格低于抛出时的价格，投资者就赚取差价。反之则会赔钱，这正是问题所在。比如，假设 1999 年 3 月 2 日，你以每股 27 美元的价格向股票投资者借得 500 股 InfoSpace 公司的股票，承诺在一年后归还，总费用为 13 400 美元。然而一年后，InfoSpace 公司的股价升至每股 482 美元，这意味着你此时必须归还的总价为 240 000 美元，几乎是当初的 20 倍。虽然最终可以证明这本该是一次聪明的选择——因为不久后 InfoSpace 公司的股价就暴降至每股 1.4 美元——但从一年期限来看，你的投资还是失败了，这也使得你将来的投资能力受

损。事实上，做空引发的损失在理论上是一个无底洞。

在实践中，借给你股票的投资者在认为你存在信用风险时，能够随时要回股票。这也意味着他能够在股价有利于他自己时选择退出交易，因为股市中存在这么个大问题，它使得被高估的股价在恢复到较为公平的价格水平前，会被更严重地高估。另外，当出借人了解到你可能必须动用存款来还清欠款时，他有权向你索取极高的利率，而你为了继续这笔交易，通常都会接受他的条件。股市泡沫要经年累月才会消退。正如约翰·梅纳德·凯恩斯所说："在市场回归理性前，你可能已经破产了。"

有时，价格的确是错误的！

但很多时候，投资者也许根本没有卖空的机会。由芝加哥大学的经济学家理查德·泰勒和欧文·拉蒙特记录的 3Com 公司事件就是一个典型案例。在这个案例中，3Com 公司决定拆分当时颇受市场欢迎的移动电话业务部 Palm，单独发行一只股票。但 3Com 公司仍旧持有绝大部分的 Palm 股票，所以投资者只要买入 3Com 公司的股票，就等于投资了 Palm 股票。公司还特别向股票持有人保证，每持两股 3Com 公司股票就能获取 3 股 Palm 股票。这似乎在暗示 Palm 股票最多能以 3Com 公司股票价值的 2/3 进行交易。

可是，当时 Palm 股票很是走俏，尽管 3Com 公司持续赢利，但是人们总觉得它不够灵活创新。结果 Palm 股价非但没有低于 3Com 公司股价，接连几个月它都以更高的价格在市场上交易。这就给了投资者一个机会，暂不论他们对 3Com 和 Palm 的评价是什么，只要通过买入 3Com 公司股票和做空 Palm 股票就保证能够获利。从理论上讲，这是一次毫无风险的套利机会，就好比是在英国伦敦希思罗机场时 1 000 美元能兑换 600 英镑，而飞机降落在纽约时 600 英镑则能兑换到 1 500 美元。

事实证明，要卖空 Palm 股票相当困难，因为几乎没有一个 Palm 股票的持有人愿意出借股份，他们都期待能从中大赚一笔：每年能有 100%的投资回报率。这一模式在互联网泡沫时期非常普遍。要想卖空网络股不是没有可能，但令人咋舌的价格确实让人望而却步。

我和泰勒在拉斯韦加斯的某次会议上都发言了，会后我们一起吃了一顿昂贵的寿司，然后一起看了拉斯韦加斯的街头表演。泰勒是法玛的同学兼同事，但他长期从事行为经济学领域的研究，也对有效市场假说发起过挑战。行为经济学指出，在所有可能的情况下，现实中的交易者的行为表现都不可能像模型中的一样好。

泰勒对我说："有效市场假说由两部分组成，其中一个部分，我称之为'没有免费的午餐'，即不可能战胜市场。对这一部分，我和尤金·法玛

观点一致。但另外一个部分，他就不怎么喜欢谈论，即‘价格是正确的’。”

有确切证据表明泰勒所说的市场中“没有免费的午餐”。确实，对一个投资者来说，长期战胜市场平均利润是很困难的（尽管不是没有可能）。理论上吸引人的机会在实际应用时可能就会变得颇具挑战性，因为受交易成本、交易风险以及其他制约因素的影响。一旦投资者发现了那些以往一直十分可靠的统计模型，这些模型就可能被证实为不可靠的。

有效市场假说的第二部分，即泰勒提到的市场的“价格是正确的”，更加不可靠。如果价格是正确的，3Com 公司股票和 Palm 股票出现价格矛盾这样的案例就能够避免。同样一款商品（同样是 Palm 股票，价值不变）却以两个不同的、严重分离的价格进行交易，这表明其中至少有一个价格是错误的。

股市中存在着很多不对称的问题，比如股市泡沫很容易觉察但很难击破。这就意味着贝叶斯定理的基本原则——若你真的认为股市将要崩盘，那为什么不放手一搏——在现实世界中是行不通的，因为现实生活中交易有限，资金有限。

金融市场中的噪声

非理性交易者和娴熟交易者之间存在共生关系，正如在扑克牌游戏中，牌桌上需要一些新手，这样出色的玩家才有利可图。在金融著作中，非理性交易者被称为“噪声交易者”。1986 年，费雪·布莱克写过一篇文章，题目就叫“噪声”：

> 噪声交易者让金融市场交易的存在变成可能，这样我们才能观察金融资产的价格。但是，噪声交易者们也导致市场无效运转……总体来看，噪声会使那些金融或经济市场运转方式的实用理论或学术理论难以检测。大多数时候，我们都在摸着石头过河。

假设市场中不存在噪声交易者，大家下赌注依据的信息都是真实的，都是信号，价格也几乎都是合理的，那么这个市场就是有效市场。

但是，如果市场是有效的——有效到你都无法战胜它而获利——那么，你作任何交易都是非理性的。事实上，有效市场假说存在着自我否定的含义。如果所有投资者都相信这个理论——市场是无法战胜的，他们就不可能从交易中获利，就没有人愿意进行交易，市场也就不存在了。

这个自我矛盾的特性，让我想起了一则关于经济学家的笑话。有一位经济学家看见路上有一张 100 美元的钞票，正打算弯腰去捡，另一个经济学家对他说：“别费力了，如果那真是 100 美元，别人肯定早就捡起来了。”如果人人都这么想，当然没有人会费力捡起那张钞票。这时一个从未学过经济学的天真小伙子走了过去，弯腰捡起了钞票，而且发现这的确是一张真钞票，于是这位小伙子用它买了一辆新车。

解开这一个矛盾的方法就是让投资者赚得一点蝇头小利，刚好能够抵消他们所付出的努力。这一行之有效的方法是多年前由诺贝尔经济学奖获得者约瑟夫·斯蒂格利茨及其同事桑福德·格罗斯曼提出的。在现实生活中，这个方法不难实现。华尔街上的证券分析师年收入可达 750 亿美元，尽管这已经让你心里有些不平衡，但相较于纽约证券交易所一年约 17 万亿美元的总交易额，就微不足道了。只要能赢取 0.5%的市场交易额，这些分析师就能维持公司的正赢利状况。

斯蒂格利茨提出的均衡手段，是让投资者得到一些蝇头小利。可有效市场假说基本上还是不成立的。尽管有些研究（包括我对亿创理财的共同基金所做的研究）似乎为法玛的观点提供了证据，证明了投资者无法战胜市场，但其他那些研究更加模棱两可，只有少数几项研究发掘出确凿的证据，证明还是有交易技术和超额利润存在的。也许战胜市场的并不是共同基金，它们依据的策略过于传统，而且投资者都成了一根绳上的蚂蚱。也许一些（并非全部）对冲基金能够打败市场，而像高盛集团这样的精英公司，它们的自营交易部门基本上都能战胜市场。另外，在期权交易者身上很容易就能发现交易技术，这些交易者主要是对股价升值空间的概率评估下赌注。然而，大多数购买量大的个人投资者即散户都会犯一些常见错误，比如交易过于频繁、赢利低于市场平均值，只有少数个人投资者才有可能击败华尔街。

直觉判断让投资者深陷“高买低卖”的误区

你不应该急于成为一名期权交易者。正如传奇投资人本杰明·格雷厄姆所指出的那样，在股市中，知之甚少危害无穷。另外别忘了，任何一个投资者其实不费吹灰之力就能做得和一般投资者一样好，他需要做的不过就是买入一只指数基金，让它随着标准普尔 500 指数的平均值而变化，只要这么做，基本上他就能复制其他所有交易者的平均利润，包括从哈佛大学投资人到噪声交易者再到乔治·索罗斯的对冲基金经理人等。若想驳倒这一命题，你要么无比优秀，想出反证的办法，要么有勇无谋，连这一命题的要求都实现不了。毕竟股市竞争激烈，尤其在今天这样一个由机构投资者主导的股市中，就连一般的股票经纪人都拥有成

堆的证书，智商极高，经验丰富。

亨利·布洛杰特对我说："人人都认为自己拥有这样一名才智过人的共同基金经理，曾求学于哈佛大学，拼杀股市 25 载。可为什么他还是无法战胜市场呢？因为股市中有 900 万名和他一样优秀的交易商，每人都有 5 000 万美元的预算，每人在纽约证券交易所都有一台计算机。在这种情况下，要如何打败由他们构成的市场？"

实际上，每天都进行投资的交易者中的大多数人的表现还达不到这么好。美国"盖洛普咨询公司"和其他一些民调机构会定期对美国民众进行调查，询问他们是否认为现在是投资股票的好时机。从历史数据来看，该民调结果和股票走势有着很密切的关系，但是负相关的关系，这与一些可靠的投资策略所显示的关系正好相反。美国民众认为，市盈率趋高且股价被高估时，恰恰是买入股票的好时机。盖洛普的记录显示，2000 年 1 月时民调结果达到了最高值，有 67%的美国民众认为这时是投资股市的好时机。然而，仅仅两个月后，纳斯达克综合指数及其他股指就开始暴跌。与此相反，1990 年 2 月，只有 26%的美国民众认为此时投资股市是正确的，结果此后的 10 年，标准普尔 500 指数几乎增加了 3 倍（图 11–10）。

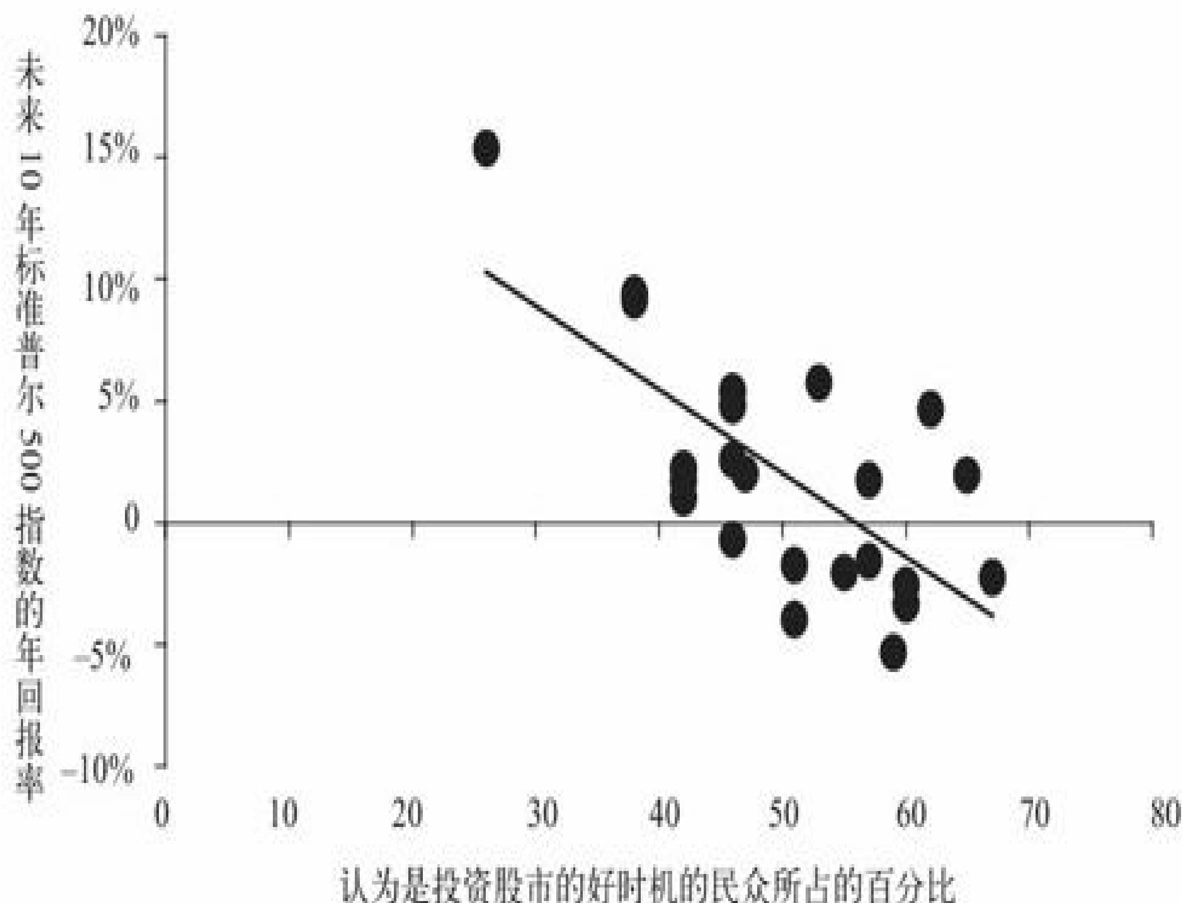


图 **11-10** 美国民众的股市投资意见和股市未来十年的年回报率

大多数人都要与自身的直觉判断作斗争。“投资者需要挣脱‘攻击或逃离’的机制的控制，做出与之相反的反应。”布洛杰特如此对我说：“当股市崩盘时，我们应该兴高采烈地投资股市，而不是被吓得急忙撤出资金。你应该明白，股市越是下跌越能产生更多利润。正常的投资者都被淹没了，因为他们在持续做错误的事情。”

正如布洛杰特所说，投资者要为这一错误付出高昂的代价。假设你在 1970 年对标准普尔 500 指数投入 10 000 美元，计划到 2009 年退休时再套现。这 40 年里股市跌宕起伏。但如果你一直都没有再买入或卖出，到退休时，你将获利 63 000 美元，这还是考虑了通胀因素、扣除本金的结果。而如果你打算“小心求稳”，多次买入、卖出，每当股价跌破上一个高峰的 25%时，就撤出资金，直到股价重回原价的 90%时才再次买进，在这样的策略下，你将获利 18 000 美元——每年仅有 2.6%的利润回报率。不幸的是，许多投资者都采取了这个策略。更糟的是，他们还倾向于在股价被高估时进行初次投资，那么，无论投资哪种股票，想要长期获得正回报，他们都得付出一番努力。

当股市泡沫再次发生时，你看到的信号会像赌场门口闪烁的彩灯一样，吸引你再靠近一点，这些信号有：消费者新闻和商业频道（CNBC）的股票价格收报机，显示出一片绿箭头……或是《华尔街日报》的头条，有关创纪录的回报……又或是发给在线经纪人的广告，宣传赚取财富只需点一下鼠标……要想避免在泡沫时期买入股票或是在经济恐慌时卖出股票，都需要保持冷静清醒、深思熟虑、镇定自若，才能避免重蹈他人的覆辙。

丹尼尔·卡尼曼将这个问题比作米勒–莱尔错觉，如图 11–11 所示，这个著名的视错觉图形由两条带箭头的线组成，它们其实一样长，但是箭头开口向外的线给人的感觉更长，而箭头开口向内的线则留给人们短一截儿的感觉。当投资者从股市中获得收益后，在他们眼里，股市就像那条箭头开口向外的线；而当投资者遭遇股市崩盘，股市就像箭头开口向内的线。

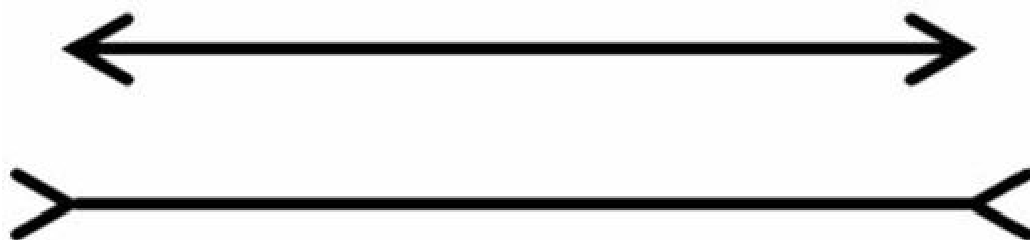


图 **11-11** 米勒–莱尔错觉

“你很难控制自己不产生这样的错觉，”卡尼曼说，“你盯着两条线看，会觉得其中一条比另一条长。但是，你可以训练自己认识到：这是一种心理模式，它会产生错觉；这时你的印象是不可靠的，应该使用规则。”

我们所依赖的认知捷径，即所掌握的策略，常常使我们这些投资者身陷麻烦。认为某个上升的事物将会持续上升，这个观点是再本能不过的反应了，但运用在股市中，这个策略就完全错了。

从众的本能可能是更根本的问题。有时，做他人所做的是完全正确的选择，或者说，他人所为至少是值得关注的。比如，到一座陌生的城市旅游，在挑选餐馆时，你可能就会选那些顾客较多的地方就餐，遇到其他状况也是一样。但是，有时这种从众策略只会适得其反，会使你陷入旅游陷阱。

贝叶斯王国中也有同样的命题，作预测时，我们理应对周围的人予以关注，并借鉴他人的理念，调整自己的观念，而不是顽固不化，死守着不太合乎情理的观念，认为自己知道的比别人多。

我作预测时便十分注意大家的共识，比如像 Intrade. com 网站这样的博彩市场里大家一致的看法。但这绝非强迫每个人都要参考他人的观点。但是，当我越是偏离一致观点时，我的论据就要越确切，直到得出结论为止。我的处理方法是对的，我认为，大多数时候这样的态度也是有助于你的。也就是说，尽管有时你能打败市场，但你不能指望每天都能打败它，如果你那样想的话，就有过度自信的嫌疑。

但是，凡事都有例外。费雪·布莱克估计 90%的股票市场都是理性的。剩下的 10%是由噪声交易者主导的市场——市场会有些失控。另外一种审视的方法认为市场通常都是基本正常的，只是偶尔出错。顺带提一下，这也是现实中的市场泡沫很难被击破的另一个原因。每 15~20 年，在你购买的资产类别里，能有一次做空泡沫或是延长恐慌的机会是非常棒的一件事。但是，想凭此维持稳定的投资生涯是十分困难的，因为你可能几年的时间都不会有收获。

有没有可能预测到市场泡沫的出现？

一些理论家建议，我们应该将股市看作一个双轨合一的车道。一条是信号轨，如我们在课本中了解到的 20 世纪 50 年代的美国股市。这类股市能长期存在，股市投资者进行的交易相对较少，股价和基本原则紧密相关，有助于投资者进行退休规划，也可以帮助公司积累资本。

另一条是快车道，即噪声轨，它所代表的股市中有很多动量交易和积极回馈，也充满了扭曲动机和从众行为。这条轨道就像一场“石头剪子布”的游戏，风险高，对更广义的经济没有什么实质性好处，但是也没有什么实质性伤害。这就是一条快车道，一群汗流浃背的投资者在上面奔跑。

然而，这两条轨道碰巧出现在一个路面上，这就好比某个城市决定举办一场一级方程式赛车比赛，但是某个政府监管机构忘记关闭通勤的车道一样。有时，路面上会发生一场大的事故，这就好比股市遭遇金融危机一样，许多普通投资者会遭到无情的碾压。

物理学家迪迪尔·索尔内特称这种二元性为“有序和无序的斗争”，这种现象在复杂的系统中很常见，这些系统受各个部分相互作用的支配。这些系统似乎同时兼具预测性和不可预测性。例如，地震系统通过一些简单的规则就能很好地描述地震（比如，我们十分清楚洛杉矶较长时间段里发生 6.5 级地震的频率）。但是，每一天的地震发生频率基本上是不可预测的。这些复杂系统的另一个特点就是，它们会发生周期性的非线性变动，这些阶段性变动十分剧烈，能从有序到混乱再到有序。索尔内特及另外一些人以高度的数学眼光看待市场，对他们而言，周期性泡

沫的出现似乎是不可避免的，是市场系统内在的特性。

我不是十分认同这个观点。我对交易市场的看法（或者说，对更广义的自由市场资本主义的看法）正如温斯顿·丘吉尔对民主的看法一样，除其他所有经济系统之外，我认为这是现今创造出的最糟糕的经济体系了。大多数时候市场的运行都良好，但是我认为我们无法就此摆脱泡沫的威胁。

如果不能阻止产生市场泡沫的羊群效应，那我们是否有办法预测泡沫的出现？假设你认同巴拉克的命题——市场在 10%的时间里是非理性的，当处于那个非理性阶段时，我们能否意识到这一点呢？因为只有意识到了，我们才能从泡沫中谋利。或者从更大公无私的角度看，我们才能制造一些软着陆以减缓某些令人生厌的纳税人的救助需求。

洞察泡沫的希望似乎没那么渺茫。虽然我认为我们无法百分之百地洞察出泡沫，甚至连 50%的概率都没有，但我们还是会有所收获的。近些年来，就有许多人提前发觉了一些泡沫，尤其是房地产泡沫。过去的经验证明，市盈率可以成为可靠的判断泡沫形成的指标。

我们也可以为问题的解决途径立法，但那样事情就变得复杂了。若某些情况需要更严格的监管，那么限制卖空制度——这会使击破泡沫变得更加困难——必然适得其反。

然而，有一点很明确，如果我们认为市场是完美的且价格永远是公正的，那就永远无法发觉泡沫的存在。市场掩盖了我们的某些缺点，均衡了我们的某些错误，确实不容易预测。但有时，价格的确是错误的。

[1]“石头剪子布”的游戏中，平衡策略指你可以随意出“石头”、“剪子”或“布”，长期来看，你肯定不会输。但不幸的是，你也不可能赢。无论游戏进行多久，这一策略带来的回报都是零。当然，如果对手的行为是可预测的，你就可以偏离这个策略，猜透他的心思。也许你的对手像巴特· 辛普森（《辛普森一家》里的主角）一样，总是出“石头”，那你就应该一直出“布”，便可以常胜不输。但这样做存在一个问题，在你努力看透对手时，你采取的策略也变得可以预测了。一旦巴特注意到你总是出“布”，他只要改出“剪子”就能赢你。股票市场大多数的技术分析策略都遵循这种猫和老鼠的动态转变，技术分析师只想要看透其他分析师。然而，当其他投资者意识到他们的交易模式，这些模式的作用就会被削弱，甚至出现自我消退的状况。结果就是，交易者之间出现大量资金流动，但是受交易成本的侵蚀，交易总量却在逐渐降低。

[2]尽管目前这些技术都已经实现，但是真的在飞机上使用，还是会让空乘人员不高兴。

[3]例如，一项常被提及的统计数据：股市每年扣除了分红和通胀后的收益会达到 7%。这只是历史平均值。可靠的股市数据最多只能追溯到约 120 年前——如果你想了解长期的收益，这些数据并不完整。统计测试表明，真实的长期效益——我们期待的未来 120 年的效益——可能会是 3%~10% 中的任意数值，而非 7%。经济学家疑惑的“股权溢价之谜”——为什么股票的收益比公债高，和风险却不成比例——的答案也许很简单，20 世纪的股票收益是有反常态的，真实的长期效益并没有 7% 那么高。

[4]《大空头》一书简体中文版已由中信出版社出版。——编者注

第十二章
温室效应：未来 10 年，全球气温会上升还是下降？

1988 年 6 月 23 日，美国国会山异常燥热。前一天下午，华盛顿国家机场的温度约为 37.8 摄氏度，初夏的气温就达到这么高的温度这几十年来还是第一次。美国国家宇航局的气候学家詹姆斯·汉森一边擦去额头上的汗珠（因为此时参议院能源委员会听证室的空调竟然不合时宜地坏了），一边告诉美国民众说，这样的高温天气可能还会持续几天，请大家作好准备。

科学家们曾经预测到温室效应会使地球升温，这一理论早已为人们接受。汉森说，温室效应开始在气温记录中制造明确无误的信号：自 20 世纪 50 年代起，全球气温已经上升了 0.4 摄氏度，而这种升温现象不可能是自然变异导致的。汉森说："气温自然出现同等上升幅度的概率仅约为 1%，因此，人们可以有 99%的信心宣称这一时期气候变暖的趋势仍在继续。"

汉森预测高温天气会频繁造访华盛顿和其他一些城市，如奥马哈，这种天气变化已"明显得人人可见了"。他建议改良气象预测模型，但气温趋势及其变化原因是显而易见的。汉森说："没时间闲聊了，有充分证据表明温室效应已经来势汹汹了。"

这场听证会距今已差不多有 25 年了，当时就"全球变暖"所提的问题，有些也可以用来对本书提到的其他领域进行提问，比如，迄今为止人们作的预测是好是坏？科学家们在哪些事情上取得了一致意见，对于哪些事情仍在争论不休？预测中的不确定性有多大，我们该作何反应？像天气系统这样的复杂事物，究竟可不可以建模研究？气候学家是否都存在过度自信的问题呢？这一问题会发生在其他领域的预测者身上吗？政治和其他反向力量会阻碍科学真相的挖掘吗？贝叶斯定理是否有助于判定争论呢？

仔细检查证据，描述出什么是对气候预测的合理怀疑，你会发现对全球变暖问题的怀疑有别于普通博文或政治主张中经常提到的内容。

温室效应真会导致全球气温变暖吗？

本书中提到了许多预测的例子，都将相关性误认为因果关系，将噪声误认为信号。1997 年之前，人们一直认为超级杯的归属与次年的股

市走向有关。然而，这两者之间其实并没有可信的因果关系，如果以此为据进行投资，你一定会输得倾家荡产，因为超级杯这一指标是假阳性的，是错误判断。

有时，假报告也会以假乱真。即使确有信号存在，过于嘈杂的数据也会将信号掩盖。举一个无可争议的例子，人们都知道热卡摄入得越多，变胖的可能性就越大。那么，这样一个基本关系一定会在统计记录中得到清晰的体现吗？

我下载了 84 个国家的肥胖率和日常热卡摄入量的数据，这些数据都是免费使用的。从这些数据看，肥胖率和日常热卡摄入量两者之间竟然没有什么关系。以肉食为主食的韩国，日常热卡摄入量约为每人每天 3 070 卡，略高于世界平均水平。然而，韩国的肥胖率却只有 3%。相反，太平洋岛国瑙鲁每人每天的热卡摄入量与韩国大致相同，但肥胖率却高达 79%。如果将这 84 个国家的情况绘制成图（图 12–1），我们会看到肥胖率与日常热卡摄入量之间的联系非常有限，以标准统计测验评判两者之间的关系得出的结论是没有“统计学意义”。

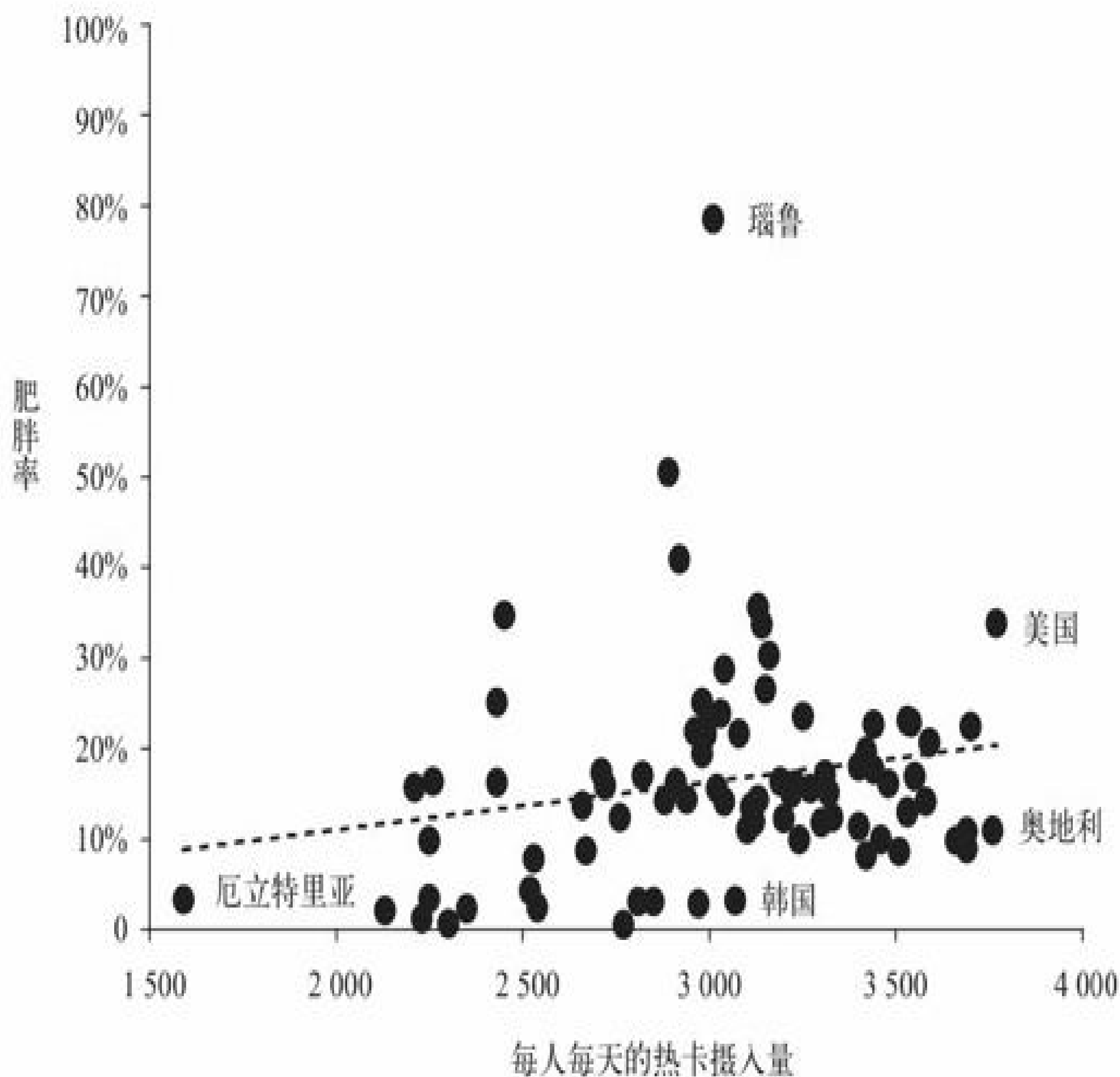


图 **12-1 84** 个国家的热卡摄入量及肥胖率

当然，掩盖两者之间关系的因素很多，如某些国家有较好的遗传基因或者锻炼身体的习惯。另外，这些统计数据十分粗糙，想要估计一个成年人的日均热卡摄入量是非常困难的。而过于看重这些统计数据的研究员会错误地否定热卡摄入量与肥胖率之间的关系，这便是错误否定，是假阴性的做法。

如果只是简单地把数据放入统计模型中进行分析，然后理所当然地认为这就是真实世界的最佳写照，事情发展若真是如此，那就好了。在某些领域中，尤其是棒球这种数据丰富的领域，提出的假设更加接近于事实真相。而在其他领域中，对因果关系疏于考虑就会走入“死胡同”。

若不是拥有因果事实作基础，人们就有理由怀疑全球变暖这一论断。在几年、几十年乃至几个世纪的时间里，全球气候经历过多次冷暖交替的变化，这种气候的不断更迭循环使得工业文明提前到来。

然而，如果现象背后存在一个对根本原因的合理理解，并以此作为预测的支撑，那么预测就更加可信。对于全球变暖的原因，我们确实有一个很好的认识，那就是温室效应。

1990 年，也就是在汉森参与那次听证会的两年之后，联合国政府间气候变化专门委员会（IPCC，以下简称"气候变化委员会"）在其第一份评估报告中公布了一项有关气候变化的调查结果，内容长达 1 000 多页。由来自全球的几百名科学家组成的团队历经数年完成了这份调查，详细地记录了气温和生态系统的潜在变化，并概述了各种可以减轻温室效应的策略。

气候变化委员会的科学家们列出了两项完全确定的调查。这两项调查不依赖复杂的模型，也没有做出与气候相关的具体预测。相反，它们依据的是相对简单的科学知识，这些知识已经为人们熟知长达 150 多年，即使是自称为气候怀疑论者的人也很少对其持有异议。这些发现迄今为止仍可算作最重要的科学论断。

气候变化委员会的第一个结论是：

> 确实存在温室效应，它使地球气温升高，而这种情况本不该出现。

温室效应是指某些大气气体——主要是水蒸气、二氧化碳、甲烷和臭氧——吸收由地球表面反射的太阳能的过程。如果不是因为这一过程，30%的太阳能会以红外线辐射的形式反射回太空，地球的气温也会比实际更低一些：平均约为零下 18 摄氏度，同火星上的气温一样。

反过来，如果大气层中这些气体越多，就会有越多的太阳能无法反射回太空，而是直接射到地球表面，气温就会升高。在金星上，大气层更厚，几乎全部都是二氧化碳，平均气温为 460 摄氏度。气温之所以这么高，部分原因是因为金星与太阳的距离近，但更多是因为温室效应。

没有什么情况表明在不久的将来，地球气候会变得同金星一样。但是，大气成分的变化必然会引起气候的变化，因为气候是十分敏感的。人类文明兴起于气温变化幅度不大的地带。世界上最寒冷的首都是蒙古首都乌兰巴托，全年平均气温约为零下 1 摄氏度。而世界上最温暖的首都大概是科威特首都科威特城，全年平均气温约为 27 摄氏度。人烟稀少的地区，冬天的气温会温暖些，夏季的气温会凉爽些，但从整个地球

范围来看，气温的冷热差距还算适度。在几乎没有大气保护的水星上，一天之内的最高气温可达 400 摄氏度，而最低温度则达到零下 200 摄氏度。

气候变化委员会的第二个结论是：随着大气中温室气体浓度的增加，温室效应会加剧，全球气温也会随之升高。

> 人类活动排放的气体是增加大气中温室气体二氧化碳、甲烷、含氯氟烃和氧化亚氮浓度的主要来源。这些气体的增加会加剧温室效应，使得地球表面的平均气温升高。而随着全球变暖，最主要的温室气体——水蒸气——也会增加，这将进一步加剧温室效应。

气候变化委员会的调查提出了几项不同的主张，每一项都值得考虑。

第一，委员会认为由于人类的活动，大气中的温室气体（如二氧化碳）的浓度正在增加，这一点显而易见。许多工业生产，尤其是使用化石燃料的产业，会产出二氧化碳这一副产品。由于二氧化碳长期存在于大气中，其浓度一直在增加：1959 年夏威夷冒纳罗亚观测站第一次监控得出的二氧化碳浓度为 315PPM，到 2011 年这一数字已增加至 390PPM。

第二，委员会认为“这些气体的增加会加剧温室效应，致使地球表面平均气温升高”，这仅是以预测的形式对第一条结论进行重申。这一个预测依据的是一些简单的化学反应，几年前就经过了实验室的证实。温室效应最早是由法国物理学家约瑟夫·傅立叶在 1824 年提出的，人们通常认为该理论是爱尔兰物理学家约翰·廷德尔在 1859 年证实的，同年，查尔斯·达尔文出版了《物种起源》。

第三，委员会认为水蒸气也会随着二氧化碳等气体的增加而增加，从而加剧温室效应，这种说法有些草率。水蒸气是温室效应最大的元凶，但二氧化碳不是。如果只是二氧化碳的浓度增加了，气温的确会升高，但与迄今为止已有的观察记录以及科学家的预测不同，气温不会升高那么多。19 世纪提出并经证明的“克劳修斯·克拉珀龙关系”这一基本热力学原则认为，大气可以在更高的气温下保留更多水蒸气。于是，随着二氧化碳和其他持久存在的温室气体的浓度增加，大气气温便会升高，水蒸气也会增加，温室效应就会成倍加剧，气温又会升高。

温室效应假说的 3 种怀疑论

只有在掌握大量论据的前提下，科学家才会断言某项假说成立。温

室效应假说就符合这一标准，所以气候变化委员会最初才会从几百份研究成果中挑出一份，也是许多科学家绝对肯定的唯一成果。温室效应背后的科学原理非常简单，所以早在 19 世纪中后期，当时电灯泡、电话和汽车刚刚问世，当然原子弹、苹果手机和航天飞机这些高端发明还没有出现，而那时人们对温室效应这一科学论断就已经十分了解了。要知道温室效应可不像火箭科学那般错综复杂，高深莫测。

的确，在气候变化委员会公布研究成果之前，瑞典化学家斯范特·阿累尼乌斯早在 1897 年就曾预言工业生产活动最终会引发全球变暖，气温升高是由于温室效应而非自然原因造成，这一点也多次得到过证明。

现在提起温室效应似乎显得有些古怪。20 世纪 80 年代，在英文书籍中，“温室效应”一词出现的频率是“全球变暖”一词的 5 倍之多。但“温室效应”一词的使用在 20 世纪 90 年代才达到顶峰，之后就日渐消退，现在的出现频率反而只有“全球变暖”一词的 1/6、“气候变化”一词的 1/10，因为“气候变化”这个词现在使用率更高。

这一（词汇使用频率的）变化是因为气候学家们试图扩大温室效应理论的影响，然而，当人们极少谈论引发这一变化的原因——温室效应——时，这种关注度的变化便不出所料地误导了人们对它的看法。

比如说，2012 年 1 月，《华尔街日报》发表了一篇社论，题为“全球变暖无须惊慌”，文中有 16 名看上去对全球变暖持怀疑态度的科学家和支持者的联合署名。《华尔街日报》还为这条社论配发了一段视频，标题下面附了这样一句话：

> 很多科学家认为二氧化碳并非全球变暖的元凶。

实际上，只有极少数科学家对“温室气体导致全球变暖”持怀疑态度。普林斯顿大学的物理学教授威廉·哈珀便是“很多科学家”中的一位，但他在社论上签了名，也是视频中的受访者之一，他在视频中用两分钟左右的时间表达了自己的观点，“大多数人和我一样，认为工业生产排放的气体会导致全球变暖”。对一些全球变暖的相关预测，哈珀表示并不认同，但他未给出自己的理由。

我并非有意暗示这一点：即使证据是自相矛盾的，你也应该盲目接受某种理论。我们可以通过人们对某种理论的预测来检验它是否正确，而气候学家做出的预测却时好时坏。气温数据非常嘈杂，全球变暖的趋势也许可以证实温室效应假说，但这一趋势也有可能是由周期性因素造成的。若全球气温走高的趋势中止，该理论就会被推翻，但也可能是因为数据过于嘈杂，掩盖了信号。

但是，即使你认为我们应该采用贝叶斯定理常用的推理手段，以概率性思维来看待几乎所有的科学假说，我们也应该对那些有极强、极明确的因果关系支撑的假说抱有更大的信心。那些看似对该理论不利的新证据虽然会使我们低估这一理论的有效性，但我们应该在自己了解（或自以为了解）的地球及气候的相关信息中权衡该理论的重要性。

健康的怀疑论应该以此为基础，对新证据的优势和该理论的整体优势进行权衡，而不是为了争出高下或意识形态方面的利益而搜集事实，因为一旦争论带有党派和政治倾向，就会变得愤世嫉俗。

2009 年 12 月，联合国在丹麦首都哥本哈根召开世界气候大会。当时正值冬至时节，日短夜长，日照时间可能只有 4 个小时，气温甚低，阵阵寒风掠过丹麦和瑞典之间狭长的厄勒海峡，从时间和地点这两方面考虑，这次会议真可谓是在最差的时间和最差的地点召开的一次全球气候会议了。

更糟糕的是，哥本哈根的啤酒贵得出奇：在丹麦，以绿色技术为支撑的基础设施建设十分发达，世界一流，这都得益于对酒类等许多产品征得的高额赋税。丹麦现在的能源消耗量同 20 世纪 60 年代末期不相上下，一部分原因是该国的环保意识增强，另一部分原因是人口较少。（与之形成对比的是美国，该国当前的能源消耗量约为 20 世纪 60 年代的两倍。）可丹麦似乎在暗示人们，高能效的未来可能是寒冷的、黑暗的、昂贵的。

于是，哥本哈根贝拉中心的气氛已经毫无悬念地超出了怀疑主义，并且更倾向于彻底的犬儒主义了。我也参加了这次会议，当时还天真地以为能够看到一场关于全球变暖的科学，严谨的辩论。而我看到的却是一场政治会议，各国的分歧不可调和。

位于太平洋上的小国图瓦卢海拔较低，是最容易受海平面上升威胁的国家之一。来自图瓦卢的与会代表游走于各个会议厅之间，大声抗议着，他们认为会议提出的目标少得可怜，根本无法减少温室气体排放。而对温室气体排放负有主要责任的几个大国却自始至终争论不休，无法达成共识。

参加会议的美国总统奥巴马没有作任何准备，他已经把其政治资本押在医疗保险改革法案和刺激计划上了。而中国、印度和巴西等国因为所处地理位置比较特殊，比美国更容易受气候变化的影响，但是这些国家却不愿意做出可能会损害其经济增长的承诺，因此也没有明确各自的立场。俄罗斯的气候寒冷，化石燃料资源储量丰富，在本次会议上成了一张百搭牌。情况与俄罗斯相似的加拿大也是一样，不大可能会努力推动任何与美国意愿相悖的议案的出台。而有些比较富裕的欧洲国家与奥

地利、日本以及非洲和太平洋地区很多较为贫穷的国家则结成了形式上的联盟。但是，在全球变暖这一问题上，虽然政治具有本土性，但科学却是无国界的。二氧化碳已经迅速包围了整个地球：一辆中国青岛的柴油货车排放的二氧化碳，最终会影响到厄瓜多尔首都基多的气候。因此，要想成功地实现减排目标，各国需要达成高度一致的意见，而不仅仅是结成几个小联盟那么简单。这样的一致意见即使不用几十年才能达成，恐怕至少也需要几年时间。

在哥本哈根会议期间，我有幸与几位科学家进行了交流，其中一位就是理查德·鲁德。这位来自北卡罗来纳州的科学家说话总是柔声细语的。鲁德曾经在美国宇航局带领科学家团队搞科研，现在在密歇根大学教授气候政策课程。

鲁德对我说："在美国宇航局，我最终认识到，所谓的'火箭科学'就是用较为简单的心理学知识解决复杂的难题。科学的部分相对简单，但其他问题，如怎样制定政策、怎样根据公共卫生做出回应等都是相对复杂的问题，因为其中没有明确的因果机制。"

我们的交谈不时地被贝拉中心的扬声器发出的通讯所打断。一位带有法国口音的女士说出了下面这段话："本次会议未能达成任何共识，本次会议讨论的各个事项暂缓决定。"而鲁德却确信，未来有关气候问题的争论中，3 种类型的怀疑论将无处不在。

第一类型怀疑论源于利己主义。仅 2011 年一年，化石燃料行业在游说活动上的花费就高达 3 亿美元（约为 5 年前的两倍）。为了本章的写作，我与几位气候学家交谈过，他们在描述化石燃料行业的这种行为时用到了"阴谋"这个词，但是如果一个基于合理的利己主义的解释是充分的，我们就没有理由断言这是一个阴谋：化石燃料公司有经济激励政策使它们能够保持现有的地位，在法律上也有《第一修正案》为其辩护。然而，我们不能误认为化石燃料公司这是在试图做出准确的预测。

第二类型怀疑论属于逆向思维。在任何有争议的辩论中，有些人会发现，与大多数人保持一致大有裨益，而一小部分人则发现自己成了大家排挤的对象。在气候科学领域，这一现象尤为明显，因为这里的数据充满了噪声，并且很难以一种切身体验的方式验证气候预测。在推崇思想独立的美国，这种现象特别普遍。鲁德对我说："不管你是观察气候、臭氧还是吸烟人群，都会发现总有那么一群人对科学运算驱动的结果持怀疑态度。"

最重要的是第三类型怀疑论，就是科学的怀疑论。鲁德说："你会发现科学界有些人极为关注科学的某个方面，从某种程度来讲，如果你真的想取得什么进展，就应该尊重这些人的观点。"

对“全球气温会持续升高”预测的批判

在气候科学领域中，健康的怀疑论主要针对的是用于预测气候变化的计算机模型。宾夕法尼亚大学沃顿商学院的教授斯科特·阿姆斯特朗就是一个怀疑论者，他也是毕生研究预测的一小部分人之一。阿姆斯特朗的著作《预测原理》堪称预测领域的权威著作。阿姆斯特朗的办公室位于宾夕法尼亚的亨茨曼大厅，我在那里拜访了他。他已有 74 岁高龄了，但留着一撮山羊胡，看上去整个人好像年轻了 15 岁。

2007 年，阿姆斯特朗向艾尔·戈尔发起挑战，想要和他打个赌。气候变化委员会曾预测全球气温会持续升高，而阿姆斯特朗则认为自己的“无变化”预测——全球气温会保持 2007 年的水平——会打败气候变化委员会的预测。戈尔从未应战，但阿姆斯特朗每月都会公布自己的结果。截止到 2012 年 1 月，阿姆斯特朗的“无变化”预测在 47 个月当中有 29 个月都更为接近实际气温，胜过专门委员会“缓慢持续升温”的预测。

阿姆斯特朗告诉我，他在本质上并不怀疑温室效应的科学性，“我认为气温确实在一点点升高，在过去 150 年里，没有人为此事争论过。”

但是，对全球变暖的多数见解，阿姆斯特朗还是颇有微词的。2007 年，在向戈尔发出挑战的同时，阿姆斯特朗与他的同事凯思腾·格林用他们所说的“审计”方法对全球变暖的预测进行审查，以便判断这些预测的真伪，尤其是气候变化委员会所作的预测是否遵从他的那些预测原则。

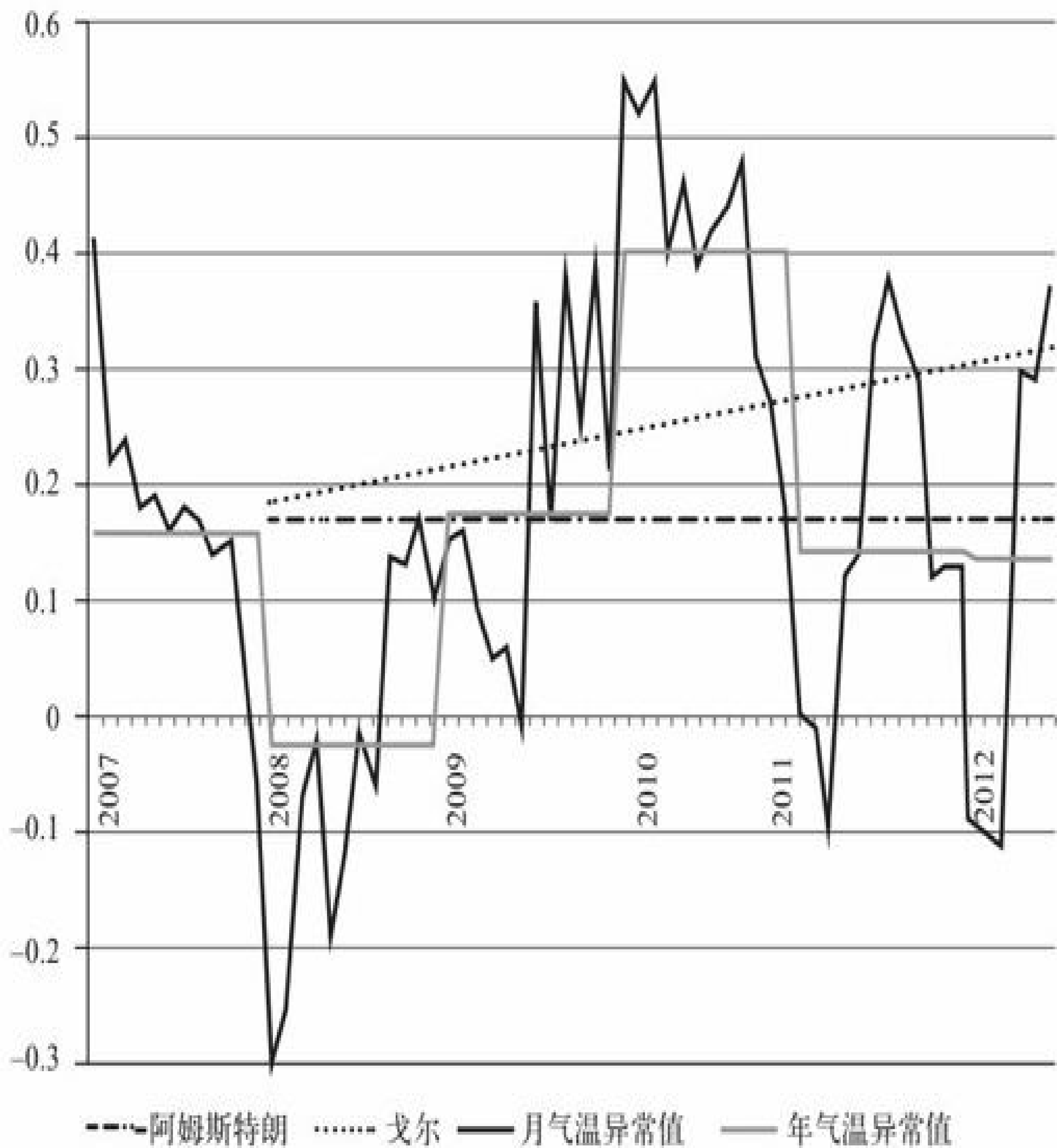


图 **12-2** 阿姆斯特朗与戈尔的打赌图示

阿姆斯特朗和格林在论文中宣称，他们发现了气候变化委员会预测的不足之处，认为气候变化委员会未能遵循 89 条预测原则中的 72 条。尽管 89 条原则可能确实显得有点多，但大部分都是适用于预测领域的实用经验，当我们用这些原则来预测全球变暖时，基本上可以把它们简化为三大批判标准。

第一，阿姆斯特朗和格林认为，预测者之间是否达成共识与其预测是否准确无关，和其他任何事一样，预测者也有个人偏见。于是，阿姆斯特朗对我说："不必投票，科学不是这样进步的。"

第二，他们认为全球变暖这一问题的复杂性，使得预测变得徒劳无功。"历史上从未出现拥有大量变量和不确定性的复杂事件最后是靠计量经济模型或其他复杂的模型解决的。模型越复杂，预测越不准确。"

第三，他们在论文中写道，预测并不足以解释全球变暖问题的内在不确定性，换言之，预测者实际上总是过于自信。

复杂性、不确定性和一致意见的重要性是本书的核心主题，每一点都值得认真倾听。

达成共识的预测结果

谈到全球变暖问题时，人们总是对"共识"一词过度痴迷，而那些不赞成共识的人也以此为荣，并把自己标榜为异端分子。另外一些人有时会依靠在线散发请愿书这类不可靠的手段，试图表达他们对全球变暖这一理论的怀疑态度。每当一些气候学家公开反对某项有关气候变暖的调查结果时，他们就会声称这些调查缺乏理论共识。

这些争论造成了人们对"共识"一词的误解。"共识"一词的内涵与"一致同意"或"多数通过"等词的内涵并不完全相同。所谓"共识"，是指组织或团体中的大多数成员围绕某一个特定的理念或方法，经过深思熟虑，最终达成广泛的一致意见。（比如，我们大多数人达成了共识，午餐吃中餐，但霍雷肖却决定午餐吃比萨。）

实际上，达成共识的过程也是一种投票方式。当某一个政党正打算选出一位美国总统提名人时，某位候选人就会在提前进行投票的州，如爱荷华州或者新罕布什尔州，作好充分的准备，因为在这些地区其他候选人都已落选。即使该候选人在得票数量上远无法达到提名的资格，但如果他向众人证实了自己已被政党中最关键的联盟接受了，也就没有必要在其他各州进行实质性的投票了。这样的候选人就是所谓的"通过达成共识赢得提名"。

确切地说，科学正是这样一种谨慎审议的过程，至少在理念上是这样的：发表文章、召开会议、验证假说、辩论结果，总会有一些科学发现能经得住考验而幸存下来。

气候变化委员会的各种活动，就是共识形成过程的示范。他们每做出一份报告就要历经数年，每一项调查结果都要经过一套全面的审核过

程，有点儿拜占庭帝国的风格和官僚主义做派。鲁德说："按照惯例，每一句审核言论都会被记录在案。哪怕是你的一位醉酒的表兄弟说了什么话，都会被记录下来。"

然而，人们对气候变化委员会这样的审议过程能做出更准确预测的期待程度，更具有争议。团队中的不同成员可以互相学习彼此的专业知识或技能，这一观念确有价值，但这可能会带来团体迷思和从众心理。有些成员更具影响力，不是因为他们有更好的想法，而是因为他们的个人魅力和社会地位。对已经达成共识的预测进行实证研究后，得到的结论各式各样，而与之不同的另一个预测过程，即由团队成员分别提交独立预测，之后经过均衡或综合的预测过程，几乎总是可以提高预测的准确度。

气候变化委员会的预测过程，可能会降低气候预测者的独立性。虽然名义上气候变化委员会有 20 个不同的气候模型可以用来作预测，但他们总是使用同样的假设、同样的计算机编码，重叠的部分太多，做出的预测只能代表五六个独立模型的预测结果。而不论模型有几个，委员会看重的只有全体成员核准的那一个预测。

不能把所有鸡蛋都放在一个篮子里

麻省理工学院的气象学者克里·伊曼纽尔是世界上研究飓风的一流理论家，他告诉我说："拥有一个多样性的模型至关重要，我们不能把所有鸡蛋都放在一个篮子里。"

伊曼纽尔说，多样性的模型之所以重要，其原因就在于这种模型除了采取不同的假设之外，还存在着不同的漏洞。"没人愿意提及漏洞，但不同的模型确实存在不同的编码错误。我们不能指望一个拥有数百万代码行和指令的模型一丁点儿错误都没有。"

如果你习惯性地认为关于全球变暖的辩论就是"怀疑论者"和"信仰者"之间的一系列争论，可能就会认为这些争论是由持怀疑论的某位科学家挑起的。实际上，尽管伊曼纽尔将自己描述成保守的共和党人——这在麻省理工学院可谓勇敢之举——但他可能并不认为自己是全球变暖的怀疑论者。相反的，伊曼纽尔在科学界声誉良好，已经被推选为美国国家科学院成员。伊曼纽尔 2006 年出版的著作中提出了一个关于气候科学的基本"共识性"（并且经过缜密思考和完善表达的）观点。

伊曼纽尔关心的问题其实在科学界相当常见：气候学家已经就某些争论不休的问题达成了较为广泛的共识。2008 年对气候学家所作的一份调查显示，几乎所有（94%）的科学家都认同气候正在发生变化这一说法，其中 84%的科学家承认"气候变化是由人类活动引起的"这一观

点。但是，对于气候计算机模型的准确度这一问题，众人则各执己见。科学家们对计算机模型预测全球气温的能力评价不一，对计算机模型模仿气候变化潜在影响的能力总体上持怀疑态度，比方说，只有 19%的科学家认为计算机模型准确地模仿了 50 年后海平面上升带来的影响。

对于任何一个对气候科学抱有讽刺观点的人而言，这样的调查结果应该都是具有挑战性的。他们总是认为科学家们在运用模型时判断不当，做出的气候预测都是捕风捉影，这样的想法应该杜绝。其实，那些科学家与他们的批评者一样，对模型也存在诸多怀疑。然而，正如阿尔·戈尔主演的那部《难以忽视的真相》一样，电影特效制作的气候变化有时不够慎重，向人们呈现了北极熊在北极为生存挣扎，或是南佛罗里达州和曼哈顿下城被洪水淹没的影像。这样的电影并不能很好地表达科学共识，气候学家争论不休的问题其实都是老生常谈，比方说如何设计计算机编码才能制作出逼真的云图？

气候科学不是“火箭科学”

气象（天气）预测者和气候学家总是意见不一，很多气象学家或含蓄或明确地批评气候科学。

气象预测者几十年来一直在努力提高预测的质量，但只要出现错误，还是会收到人们发来的表达愤怒的电子邮件。提前 24 个小时作预测确实充满挑战，所以，那些运用类比技术作预测的气候预测者怎么可能现在就预测出几十年后的气候呢？

正如谈到“共识”这个词时会出现几个意义相近的词一样，“气象预测”和“气候预测”这两个概念也存在一定的语义差别。气候是指地球自身的长期均衡状态，而气象描述的则是背离了长期均衡状态的短期表征。气候预测者不会试图预测 2062 年 11 月 22 日塔尔萨是否会下雨，他们更关注的可能是未来整个北半球的平均降雨量是否会比现在多。

而气象学家需要全力应对复杂性：混沌理论的整个学科都是从基本无望的气象预测发展而来的。气候学家也需要应对复杂性，比方说，云属于小规模现象，需要大量的计算机处理才能模仿得准确，却能对气候预测内在的反馈环产生潜在的深远影响。

充满讽刺意味的是，气象预测竟然是本书的成功案例之一。通过各种努力，结合计算机的力量和人类判断，气象预测比 10 年前甚至 20 年前准确多了。大多数领域的预测者都有过度自信的倾向，而气象预测者严于律己且严于律人的精神实在可贵。但是，他们取得的进步有力地驳斥了“复杂性使进步无望”的观点。

气象预测取得的进步源于该学科的两大特征。第一，因为每天都要

进行气象预测，所以气象学家可以获得大量反馈，可以实时检查以校准他们的预测。而气候学家在作预测时就不具备这一优势，他们往往要预测 80~100 年后的气象状况，所以人们才会对他们的预测持怀疑态度。

第二，对天气系统的物理现象（这些物理现象受相对简单的、容易观测的规律支配）有深刻了解，同样会惠及众多气象学家。其实，气候学家可能也拥有同样的优势。我们能够观察云，也能准确推断这些云的运动方式，但如何将这一推断转换为数学术语是极大的挑战。

气候预测的一个经典案例，是对某些相当重要的巨大云层的运动轨迹进行的成功预测，这些云层最终演变为飓风。伊曼纽尔位于麻省理工学院的办公室编号为 54—1814 房间，想找到这里并非易事（我在一位很特别的看门人的帮助下才找到这间办公室，这位看门人可能跟我一样，都受到《心灵捕手》这部电影的鼓舞）。从伊曼纽尔的办公室里，可以清楚地看到查尔斯河的全景，也很容易想象到远处飓风的走向：飓风会向剑桥地区倾斜呢？还是会刮到北大西洋呢？

伊曼纽尔对两种飓风预测进行了区分。一种是单纯的统计区分，“对自己感兴趣的现象进行长期跟踪记录，对自己认为可行的预测指标进行长期跟踪记录，比如在大气中大规模流动的云或海洋温度等所有可获得的信息。不必特意亲身接触这些物理现象，只需将统计数据与你想要预测的事物联系起来即可”。

设想一股飓风正盘踞在墨西哥湾上空。我们可以针对过去出现过的飓风建立一个数据库，观察其风速、经纬度、海洋温度等数据，找出与这次飓风最为相近的地方。其他飓风是怎样运动的？什么样的飓风会袭击新奥尔良这样人口聚集的地区？什么样的飓风会最终散去？不用依靠所有气象知识，只需一个完善的数据库，我们就可以做出预测。

这样的统计技术可以提供粗糙但有用的预测数据。实际上倒退 30 年，当时气象预报预测飓风轨迹的最主要的方法就是纯统计模型。

然而，此类技术会受到收益递减规律的支配。在美国，飓风并不少见，风暴平均每年都会袭击美国一次。当你将大量的候选变量应用到一个罕见的现象中，就会有过度拟合的危险，会把过去数据中的噪声误当作信号。

当然，如果你对该系统的结构有所了解，那就还有另外一个方案可供选择。第二种气候预测模型从本质上来讲，就是模拟某个领域中某些部分的物理力学特性。与纯统计模型相比，建立这种模型需要更大的工作量，并且要对该现象发生的根本原因有深刻的了解。但是，这种模型更加准确。这样的模型现在正用于预测飓风路径，而且相当成功。正如前文中提到的那样，20 世纪 80 年代以来，飓风路径预测的准确性已提

高了 3 倍左右，人们提前 48 个小时就能了解到袭击新奥尔良的卡特里娜飓风的确切登陆位置（虽然不是所有人都相信这一预测）。统计学驱动的系统现在不过是作为基准，用来对这些更加准确的预测进行衡量。

模型越复杂，预测越糟糕

阿姆斯特朗和格林对气候预测提出的批评，与他们所作的研究有关。这两位学者对经济学这类几乎不存在可用的物理模型、人们对其因果关系也知之甚少的学科进行过实证研究。野心过于膨胀的预测方法在这些领域频频失败，所以阿姆斯特朗和格林推断，这些方法用于气候预测也会失败。

任何预测模型的目标都是尽可能地“抓住信号、扫除噪声”。保持两者的平衡有时并非易事，需要有理论依据和保质保量的数据做保证。在经济预测中，数据贫乏，理论研究薄弱，所以阿姆斯特朗才会认为“（经济）模型越复杂，预测越糟糕”。

在气候预测中，情况更加模糊不清：温室效应理论的影响很大，可以支撑更为复杂的预测模型。然而，气温数据非常嘈杂，总是与预测模型相向而行。哪种考虑才是对的呢？我们可以依据经验处理这个问题，对气候科学中曾经使用过的各种预测方法的成功和失败的状况进行评估。像往常一样，最重要的是看这些预测在现实生活中的效果。

我敦促大家，不要将预测过程缩减为一系列的“车贴标签”式的标语口号。简单性原理（“在其他条件相同的情况下，简单的解释比复杂的解释更好。”）这类启发法看似诱人，却难以为我们所用。与用于预测疾病爆发的 SIR 模式一样，有些预测模型中的假设既简单又简洁，这样的情况我们已经遇到过很多，但是这样的模型却过于天真，无法提供成熟的预测。而在地震预测中，那些复杂得离谱的预测方案在软件程序包里功能齐全，在实际应用中却漏洞百出，真可谓华而不实。

“模型越复杂，预测越糟糕”，这句话就是在告诫人们，“不要在食谱中加入过多的盐”。你在做事之前是不是把它搞得很复杂，或者说在其中加了很多“盐”呢？如果你想越来越擅长作预测，就要“相信自己的厨艺，相信自己的味蕾”。

气候预测中的 3 类不确定性

认识到预测中的局限性只是成功的一半，在这一点上，气候预测者已经做得相当好了。气候学家敏锐地意识到不确定性的存在：在政府间气候变化专门委员会于 1990 年公布的 3 份报告中，“不确定”和“不确定性”这两个术语出现了 159 次。政府间气候变化专门委员会建立了一整套命名法，用于表达对某一研究发现的赞成或确定程度。比如，在报告

中，当“可能”这个词单独出现时，意味着某一预测实现的概率为 66%，而当“几乎确定”这个词出现时，意味着对某一个预测怀有 99%或更强的信心。

然而，警惕不确定性是一回事，适当准确地估计到不确定性又是另一回事。像政治民调这类问题，我们可以依靠强大的历史证据数据库：美国总统大选前一个月，如果一名候选人在民调中领先 10 个点，那么他最终赢得美国总统大选的概率是多少？我们可以查阅过去几十场美国总统大选的情况，进而依据经验得出一个答案。

但是，气候预测者建立的那些模型不能依靠此类技术。人类只有一个地球，我们要不时地对其气候演变进行预测，这种预测需要跨越未来几十年的时间。尽管气候学家也许已经考虑到预测中会存在不确定性，却仍然不能确定预测中究竟有多少不确定性。对于任何学科的预测者而言，这样的问题都是极具挑战性的。

尽管如此，我们还是有可能将气候预测中的不确定性分为 3 个组成部分。加文·施密特是汉森的同事，与汉森同在宇航局工作，他是一个说话带刺儿的伦敦人，也是实地天气网站的博文合著者，施密特的办公室位于纽约晨边高地，我是在他办公室附近的酒吧与他碰面的。

施密特拿出一张餐巾纸，在上面画出一张如图 12–3 所示的图表，说明了气候学家面临的三大突出问题。在气候预测过程中，这 3 种不同类型的不确定性不同程度地存在着。

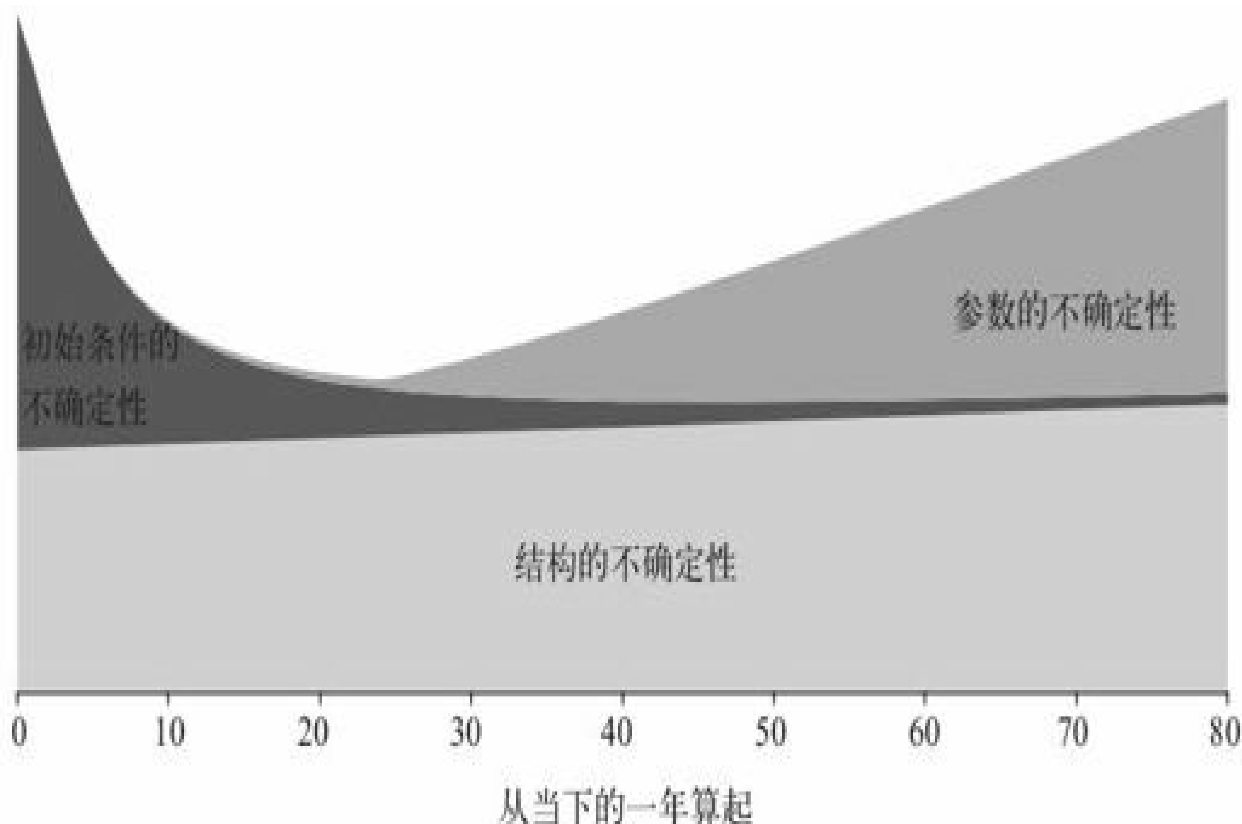


图 **12-3** 全球变暖预测中的不确定图示

第一种类型，施密特称之为“初始条件的不确定性”，这是与温室信号抗争的短期因素，影响着我们对气候的体验。温室效应是一种长期现象，会被日复一日、年复一年的各种事件所掩盖，从而变得模糊不清。

初始条件的不确定性的最为明显的例子就是天气，如果用于气候预测，那它代表的只是噪声，而不是信号。气候变化委员会最近预测，下个世纪的气温可能会升高 2 摄氏度，相当于每 10 年升高 0.2 摄氏度，或每年升高 0.02 摄氏度。可是，在温带地区，当昼夜温差达到 15 摄氏度、季节温差达到 30 摄氏度时，这样的信号是很难察觉的。

我是在 2011 年拜访施密特的。事实上，就在我们会面的前几天，纽约和美国东北部其他几个地区就出现了反常的 10 月暴雪。中央公园的积雪达到 33.2 毫米，创下了 10 月降雪的纪录，而在康涅狄格州、新泽西州和马萨诸塞州情况更加严重，几百万居民的家中同时断电。

中央公园正好有一个相当完整的气温纪录，可以追溯到 1869 年。在图 12–4 中，我绘制了跨越 1912~2011 年整整 100 年的月平均气温，从图中可以观察到四季气温的剧烈波动（变化虽然剧烈，但也能够预见），从高温到低温再回到高温，有些年份较高，有些年份较低。从这些例子中可以看出，与天气相比，气候信号几乎无法察觉，但这一信号

确实是存在的：中央公园在 100 年的时间里平均气温降低了约 15.6 摄氏度。

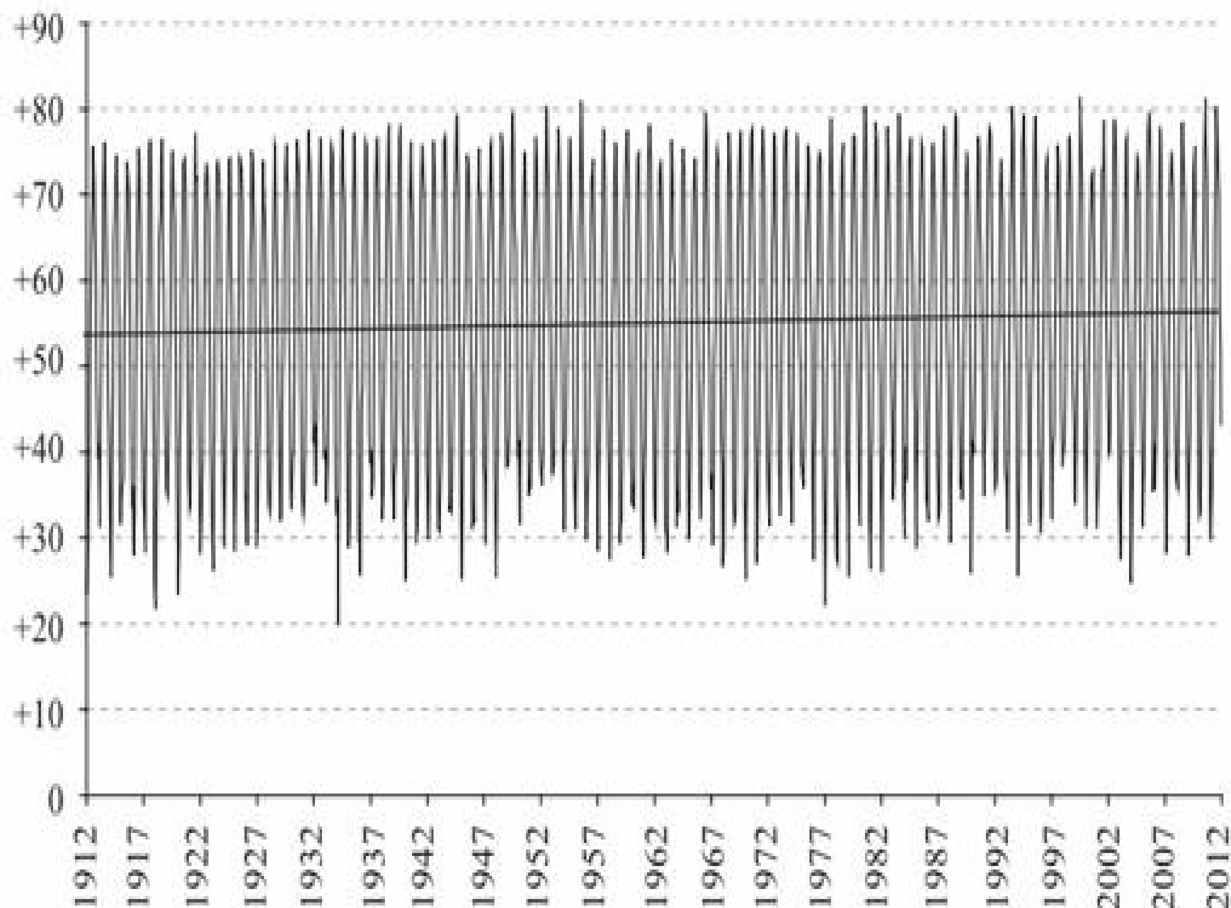


图 **12-4 1912-2012** 年中央公园（纽约市）的月平均气温

而在为期 1~10 年的时间里，气温也会出现周期性波动。其中之一被称为“ENSO 循环”（厄尔尼诺–南方涛动现象）。每次循环会历经 3 年左右的时间，是由热带太平洋地区的水温变化引起的。在发生厄尔尼诺现象的年份，若该循环充分发挥威力，就会给北半球大部分地区带来更加温暖的天气，也会降低墨西哥湾飓风的活动频率。在发生拉尼娜现象的年份，情况正好相反，这时的太平洋水温较低。除此之外，人们对 ENSO 循环的了解相对较少。

初始条件的不确定性的第二个案例是太阳的活动周期，平均约为 11 年，在这一个周期里，太阳会发出略强或略弱的辐射。（这一点通常可以通过太阳黑子这个更高一级的太阳活动来衡量。）但太阳的活动周期不怎么规律，比如，太阳的第二十四个活动周期中，人们预期太阳黑子

在 2012 年或 2013 年的活动量最大（因此会出现更高的气温），但实际上这一现象却推迟了。太阳的休眠期有时可以持续几十年，17 世纪末 18 世纪初的蒙德极小期持续了近 70 年，这段时期内几乎没有太阳黑子活动，因此欧洲和北美地区的温度大幅走低。

最后一个案例是向大气喷发硫黄的火山的周期性休眠。火山喷发释放的气体具有反温室效应的作用，可以使地球降温。1991 年皮纳图博火山爆发，此后两年里，全球温度降低了约 0.2 摄氏度，与温室效应在 10 年里带来的温度升高幅度相当。

时间跨度越大，就越容易忽略这些中期效应。在 1~10 年的时间跨度内，中期效应都能够主导温室效应的信号，但超过这一时长，它们就会日渐式微。

第二种不确定性会随着时间的推移而增加，施密特称之为参数的不确定性。这一类型的不确定性涉及大气中二氧化碳或其他温室气体的浓度。在较小的时间跨度内，大气成分还是可预测的。工业活动持续不断，二氧化碳迅速进入大气层并长期滞留。（据估计，二氧化碳的化学反应半衰期约为 30 年。）即使主要工业国家同意立即制定出实质性的减排措施，想要使大气中的二氧化碳的增长速度放缓也需要数年的时间，更不要说实现负增长了。施密特对我说："你我永远都无法等到二氧化碳浓度降低的那一天，就连你的下一代也等不到。"

由于气候模型依赖于对大气中二氧化碳浓度的特定假设，这不仅使未来 50 年或 100 年的预测变得非常复杂，也会对近期的预测产生轻微的影响，影响程度主要取决于影响二氧化碳排放的政治和经济决策。

第三种不确定性是结构的不确定性。这种不确定性是气候学家及其批评者最为担忧的不确定性类型，因为该类型最难定量，与我们对气候系统动力学的了解程度及我们能否很好地用数学方法表现这种模型紧密相关。结构不确定性的增强速度较为缓慢，而在气候这样的动态系统模型中，错误也会自我强化。

施密特认为，综合 3 种不确定性因素考虑，在气候预测出现之前的 20~25 年里，这 3 种类型的不确定性很可能都处于最低点。此时，我们就可以比较合理地确定大气层中二氧化碳的浓度了，但是考虑到 ENSO 循环、火山周期性休眠和太阳活动周期对气候的影响已经日趋平稳，我们要做的还有很多。

气候变化委员会在 1990 年公布了第一份报告，当时正值 20 年里的最佳时机。20 世纪 80 年代，詹姆斯·汉森所作的一些早期预测也是如此。换句话说，是时候评判预测的准确性了。它们究竟表现如何呢？

评判气候预测准确性的时刻到了

衡量一个预测的准确性，首先需要一个测量工具，而气候学家的测量工具很多。全世界主要有 4 个评估全球气温的组织，它们使用的测量工具从陆地上的温度计到海洋中的测量站，这 4 个组织分别是美国宇航局、美国海洋和大气局，还有英国和日本的气象部门。

最新加入这场“预测竞赛”的是卫星观测。最常用的卫星记录来自亨茨维尔的阿拉巴马大学和一家名为“遥感系统”的私有公司。卫星记录并不直接被采纳为气温数据，而是通过对微波辐射的测量进行气温预测。卫星测量出的低层大气的气温代表一个合理的表层气温。

根据记录时间长短的不同，气温记录有所不同，最早的记录可以追溯到 1850 年英国气象局的观测报告，最近的记录则是 1979 年的卫星记录。这些记录有不同的测量基准，比如，美国宇航局的记录是相对于 1951~1980 年这 30 年间的平均气温得出的，而海洋和大气局的记录是相对于整个 20 世纪的平均气温得出的。这两个记录也很容易得到校正，因为各个系统的目标是测量气温上升或下降了多少，而不是测量气温的实际数值。

令人欣慰的是，各种记录之间的差异并不大。图 12–5 中的气温记录显示，1998 年和 2010 年是气温最高的两个年份，同时还可以看出，自 20 世纪 50 年代以来，大气中二氧化碳的浓度快速升高，温度呈明显上升趋势。

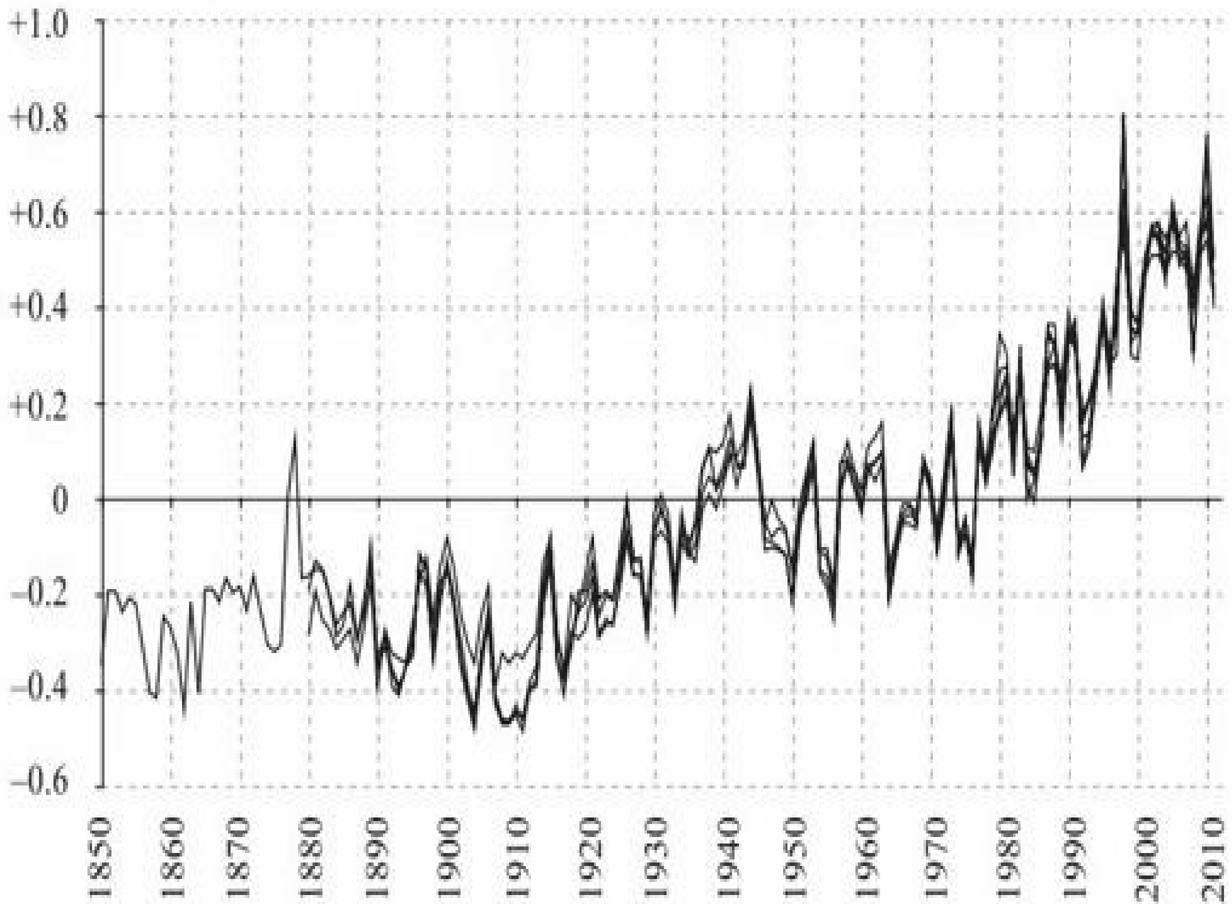


图 **12-5** 以 **1951-1980** 年的平均气温为基准的全球气温分布图

1981 年，汉森与其他 6 位科学家在学术声望极高的《科学》杂志上发表了一篇论文，预测气温的早期尝试自此开始。这些预测并不是以完全成熟的仿真模型为基础，而是依据二氧化碳及其他大气气体的影响做出的较为简单的统计预测，效果非常好。但实际上，他们略微低估了 2011 年的全球变暖程度。

汉森之所以小有名气，是因为他在 1988 年所作的国会证词以及同年发表在《地球物理学研究杂志》上的一篇论文。这些预测确实是依靠一个大气的三维物理模型做出的。

汉森在美国国会上说，华盛顿未来将会经历更多的“炎热夏季”。在其论文中，他为“炎热夏季”下了定义：所谓“炎热夏季”，是指华盛顿夏季的平均气温将达到 1950~1980 年这 30 年间夏季气温排名前十的水平。到 20 世纪 90 年代，华盛顿夏季的高温时间将达到 55%~70%，约为 33%这一基准比例的两倍。

汉森的预测最后都应验了。20 世纪 90 年代，华盛顿经历的 10 个夏

天里有 6 个的气温都达到了“炎热”的程度（表 12–1），基本符合汉森的预测。21 世纪的第一个 10 年里，这一趋势仍在继续，到 2012 年气温更是创下了新高。

表 **12–1** 炎热夏季

基准比例：33% 汉森的预测（1988 年）：20 世纪 90 年代华盛顿夏季的高温时间将达到 55%~70%				
城市	临界值（华氏度） 1 华氏度 =–17.2 摄氏度	1990~1999 年	2000~2011 年	1990~2011 年
华盛顿	86.2	60%	58%	59%
奥马哈	86.2	10%	42%	27%
纽约	81.4	80%	75%	77%
孟菲斯	89.3	50%	67%	59%
平均值		50%	61%	56%

汉森在其论文中还为其他 3 个城市做了预测：奥马哈、孟菲斯和纽约。得到的结果更加多样，体现了气候的区域差异。根据汉森对“炎热夏季”的定义，20 世纪 90 年代，奥马哈只有一年的夏季达到“炎热”的程度，低于 33%的历史平均比例。而根据拉瓜迪亚机场的观测，纽约的 10 个夏季里有 8 个达到了“炎热”的程度。

对 4 个城市的预测总体来说相当不错，但只能算是汉森的低端预测。汉森的全球气温预测更难评判，因为这一预测涉及的场景太多，各自依赖的假设也不同，实属高端预测。即使是最保守的方案，也多多少少地高估了 2011 年经历的气候变暖的过程。

作为国际社会在这一领域做出一致性预测的首次尝试，气候变化委员会于 1990 年的预测受到了广泛的关注。虽然与汉森的预测相比，这次预测还是不够明确，但是他们的预测详细周密，较为准确。比如，他们预测陆地表面与海洋（湖泊、河流）表面相比升温更快，而且在冬季更为明显，另外，在北极和其他北半球高纬度地区，将出现气温大幅度升高的现象。结果证明，这两个预测都是正确的。

然而，政府间气候变化专门委员会所作的头条预测是“全球气温将会升高”，为自己的预测留下了很多悬念。

与汉森不同，委员会的气温预测采取的是呈现一系列可能的结果的形式。在所有可能的情况中，最高值是下一个 100 年里气温将会灾难性地升高 5 摄氏度，最低值是下个世纪的气温只会升高 2 摄氏度，而更有可能出现的情况是全球气温升高 3 摄氏度。

实际上，自从这份报告公布以后，实际气温的上升速度已经减缓了（图 12–6）。在 1990~2011 年间，每年气温平均升高 0.015 摄氏度，或每 100 年升高 1.5 摄氏度，这只是委员会预测报告中提出的最有可能的数值的 1/2，也稍低于最低值。另外，委员会的预测报告还高估了海平面上升的幅度。

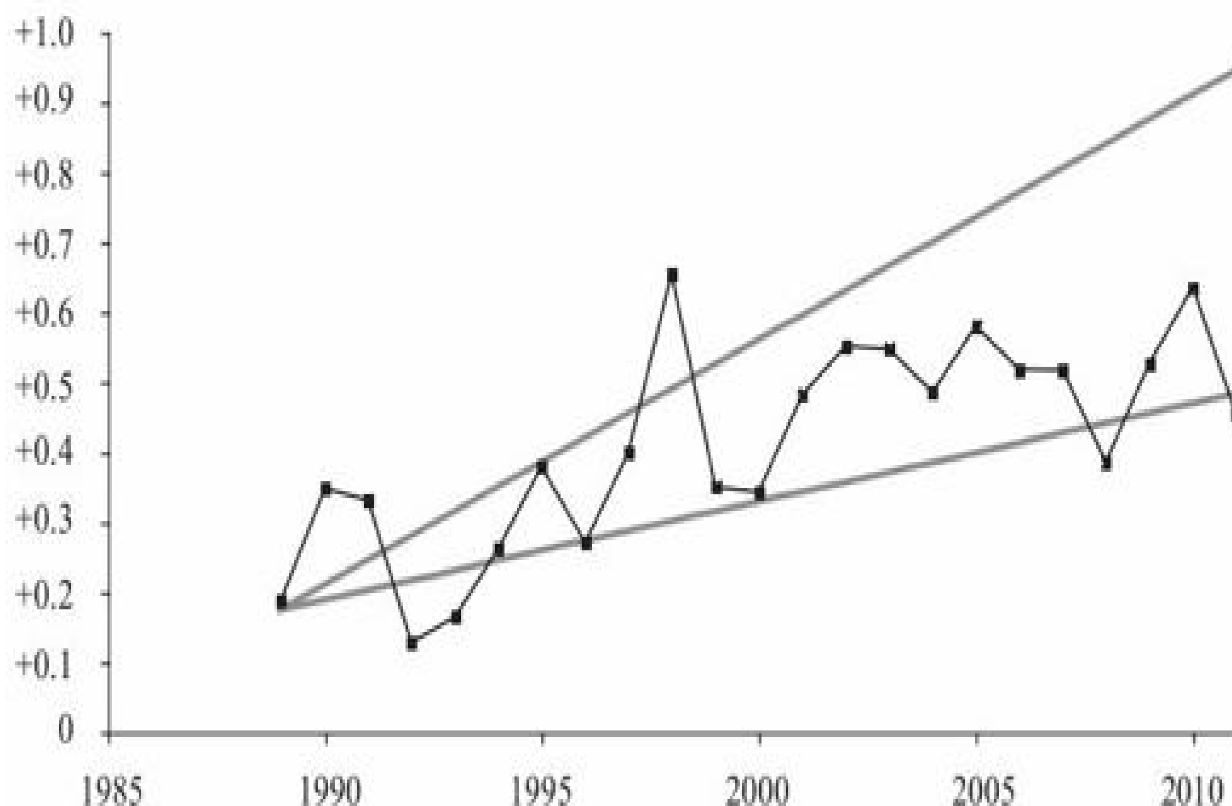


图 **12-6 1990-2011** 年全球实际气温 **VS.1990** 年气候委员会的预测值

这样的结果是对气候变化委员会预测效果的一次沉重打击，尽管我们应该考虑一个重要的限制条件。

政府间气候变化专门委员会是在“一切照常”的情况下做出的预测，

认为减少碳排放量是不可能做到的事。这套方案意味着，截止到 2010年，大气中的二氧化碳浓度将会达到 400PPM。实际上，人们已经在为减排努力了，尤其是欧盟，虽然他们付出的努力十分有限，但是这一预测也太过悲观了，2010 年二氧化碳浓度实际上仅为 390PPM。换句话说，预测中的错误其实反映了预测中参数的不确定性——这对政治和经济问题的影响比对科学问题的影响更大，还反映了委员会对减排努力过于悲观的态度。

尽管如此，委员会最终还是承认其预测过于冒进。1995 年，委员会成员公布了另一个预测，“一切照常”情况下的预测范围经过修正，使得预测值大大降低：每个世纪气温升高 1.8 摄氏度。与实际的气温趋势相对比，这次预测已经算是非常准确的了，它基于委员会做出的巨大调整。因此，当你发现预测错误时，正确的做法是及时改正，而不是学堂吉诃德的“愚侠”精神，誓死不屈。这也从侧面证实了气候预测中确实存在不确定性。

对早期预测的全部努力给出的分数，可能取决于你是否根据曲线图打分。气候变化委员会于 1990 年犯下的预测错误在一定程度上解释了预测中存在参数的不确定性，但如果委员会在 5 年后做出的预测仍然没有实质性改变，这一论点就更有说服力了。另外，1995 年的气温预测基本准确，为数不多的几个关于全球气温升高（如北极冰面收缩）的预测也相当不错。如果以高标准评判预测者，那么委员会的分数就会比较低，但也能勉强及格。如果你已经了解到预测史原本就充满了失败，那么相比之下，委员会的预测已经相当好了。

预测中的不确定性未必能成为不采取行动的理由，相反，耶鲁大学经济学家威廉·诺德豪斯认为，正是由于气候预测中存在不确定性，人们才不得不行动起来，这是因为气温升高是一件糟糕的事。另外，在气候科学中，预测更具有猜测性而不具有准确性，政府正是打着这个幌子，要么在经济刺激计划上投入数千亿美元资金，要么发动中东战争。

“全球变冷”事件的教训

时至今日，气候学家的每一次预测都冒着声誉受损的危险：其他领域的错误预测很快就会被遗忘，而相比之下气候预测的错误却会被人们牢记数十年。

气候评论家曾经普遍认为，人类会遭遇全球变冷现象和新的冰河世纪。确实，20 世纪 70 年代发表的几篇论文曾经预测全球有变冷的趋势。这些论文以一个足够合理的理论为依托：由硫排放引起的变冷趋势将压过由碳排放导致的变暖趋势。

这些预测遭到了主流科学文献的驳斥，但得到了新闻媒体的支持。1975 年，《新闻周刊》上的一则报道曾经猜想泰晤士河和哈德逊河可能会结冰，并且认为食品生产会大幅度减产，不过这些观点仅代表文章作者本人的观点，与所有接受采访的科学家的看法无关。

如果媒体在气候科学争论者中的“怀疑论者”和“信仰者”之间错误地画上等号，那它们有时也能舍本逐末地选出最荒唐的气候变化断言，即使这些断言受到一大群科学家的批判也无所谓。

2011 年纽约下起 10 月怪雪之后，施密特曾这样对我说过，“问题是，很多人会到处乱讲，好像他们亲眼见过这些数据似的，我敢保证谁都没有亲眼看到数据”，当时许多媒体都把这场雪灾视作与全球变暖相符或者相悖的证据。

那时施密特也接到了无数通电话，很多记者都在问他这场纽约 10 月暴雪在气候变暖方面意味着什么。施密特告诉这些记者他本人也不能确定，模型并不能计算得那么具体详细。但是，他的一些同事的言论就没有那么谨慎了，越是引人注目的言论就越有可能被新闻媒体引用。

20 世纪 70 年代，人们依据硫排放问题做出全球变冷的预测。硫排放也许可以解释气候变化委员会在 20 世纪 90 年代的预测为何会失败，也可以解释 1995 年专家组为何大幅缩小气温预测的变化范围。1991 年，菲律宾吕宋岛的皮纳图博火山爆发，向大气中排放了含硫气体，这次火山喷发带来的影响和气候模型的结果是一致的。但是仍需强调的是，不同温室气体的相互作用会给气候模型带来挑战，还会导致系统误差。

从图 12–7 可以看出，人为因素导致的硫排放在 20 世纪 70 年代早期达到峰值，随后开始下降，这在一定程度上是由于当时尼克松总统为了解决酸雨和空气污染问题，于 1970 年颁布了《清洁空气法》。20 世纪八九十年代的全球变暖趋势或许可以反映出硫排放量降低的走势，因为二氧化硫的排放能够抵消温室效应。

然而，大约从 2000 年以来，硫排放量再次增加，主要是由于发展中国家的工业活动越来越频繁，大量燃煤电厂投入生产。尽管硫排放对全球变暖趋势的削弱作用并不像碳排放量增加产生的影响那么大——否则，全球变冷理论就是正确的了——但是，硫排放或许真的可以阻止全球变暖的进程。



图 **12-7 1900-2005** 年全球硫排放量

正确的预测绝对离不开科学的方法

如果某一个预测声称能够对气候这类复杂的过程做出准确的预测，或者称需要几年时间才可以证明其准确性，那么人们必然会对其持怀疑态度。

一知半解的预测者有时会错误地认为，他们之所以忽略某件事，是因为对这件事建模是一件难事。优秀的预测者总是有备用计划，如果他们发现自己的模型是失败的，就会默认使用一个合理的基准预测。（在美国总统选举中，你的默认预测是在任的候选人会获胜，这比在候选人中随机挑出一个要准确得多。）

那么，气候预测的基准方案是什么呢？如果人们批判全球变暖的预测不切实际地复杂，就要有一个相对简单的预测取而代之，这个预测需要有强大的理论假设，但没有其他不必要的附属条件。

比方说，假设你试图以一个极其简单的统计模型为基础作气候预测，这个模型忽略了硫排放、ENSO 循环、太阳黑子等各项参数，仅根据二氧化碳浓度、气温等变量进行预测。此时不必使用超级计算机，只需一台手提电脑在几微秒内就能够计算出结果，但这样的预测准确性能有多少呢？

实际上，与气候变化委员会所做的预测相比，这样的预测很可能非常准确。如果将 1850~1989 年的气温记录、南极冰芯及位于夏威夷冒纳罗亚观测站测量出的二氧化碳浓度都放入一个线性回归方程式中，得出的结果是，从 1990 年至今，全球气温正在以每个世纪 1.5 摄氏度的幅度上升，这与图 12–8 所示基本一致。

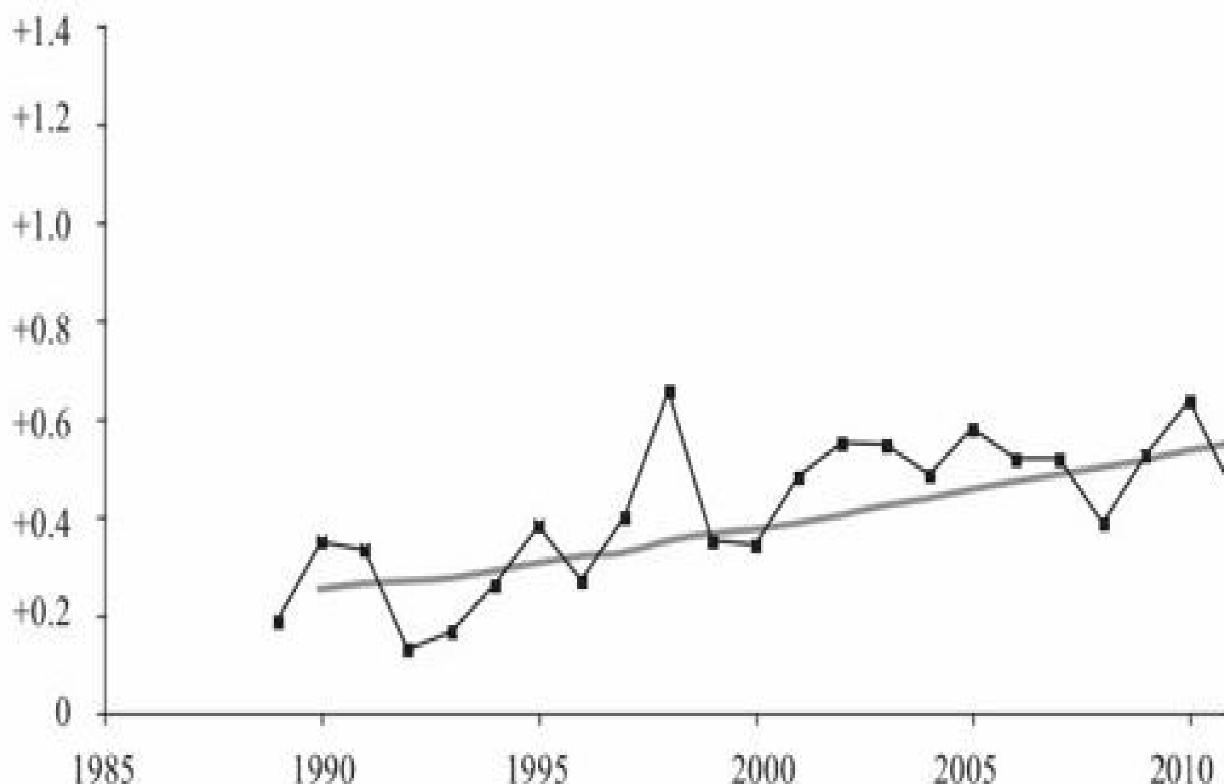


图 **12-8 1990-2011** 年全球实际气温 **VS.**简单回归预测

另外一种方法只是稍显复杂，即使用当时十分容易获得的有关二氧化碳浓度与气温之间总体关系的预测值。对全球变暖的所有预测有一种“通用货币”，即大气中二氧化碳浓度加倍（即 100%增长）对气温产生的影响的值。长期以来圈内人在一定程度上对这个值的认定是一致的。从 1938 年英国工程师 G·S·卡伦德依据简单的化学方程式做出的预测到当今超级计算机做出的预测，从所有预测值中总结出的结论是：二氧化碳增倍会使气温上升 2~3 摄氏度。

考虑到大气中二氧化碳的实际增长速度，通过简单转换得出的结果也许已经表明，从 1990 年至今，气温上升速度为每个世纪 1.1~1.7 摄氏度。实际上，气温上升速度为每年 0.015 摄氏度，即每个世纪上升 1.5

摄氏度，这个数值与预测值正好吻合。

詹姆斯·汉森 1981 年的预测所使用的方法与上述方法十分相似，在对当时的气温进行预测时，效果相当不错，比他在 1988 年依靠模拟气候模型做出的预测要准确得多。

由此看来，阿姆斯特朗和格林对复杂模型的批判是对的。比较基础的预测方法虽然成功了，但也只能表明阿姆斯特朗赢得的只是一场战斗，而不是整个战争。阿姆斯特朗对复杂模型提出了质疑，而简单模型能做出更佳气候预测的事实也支持了他的立场，所以他认为简单模型更胜一筹。事实上，简单模型的确正确地预测出气温会随着二氧化碳浓度的增加而上升，同时也证实了温室效应的假说。

与此不同的是，阿姆斯特朗的“无变化”预测留下了一些最基本的科学问题等着人们去解答。2007 年的气温虽不是高得出奇，但却高于 20 世纪任何一年的气温，而“无变化”预测正是以 2007 年的气温为基准。对于 2007 年的气温高于 1987 年、1947 年或是 1907 年这一事实，除了大气成分的变化之外，还有什么合理的假说可以对此做出解释呢？实际上，气候模型最切实的贡献，就是发现除非我们能够对大气中二氧化碳和其他温室气体浓度升高的原因做出解释，否则想复制目前的气候基本上是不可能的。

阿姆斯特朗对我说，他之所以做出“无变化”预测，是因为他并不认为贝叶斯定理优于其他任何假设，而在其他研究领域，“无变化”预测是一个非常好的默认预测方式。如果将气候预测的严谨做法运用到其他研究领域，将更有说服力。但是阿姆斯特朗并没有这么做，正如 2011 年他对美国国会监督小组所说的那样，“我确实在努力学着不那么关心气候变化，可我就是一个喜欢预测的人，没办法。”

有些预测者说科学对预测来说并不重要，而有些科学家则说预测对科学来说并不重要，本书就是想提醒你，对这两种人的观点皆不可轻信。科学与预测在本质上是紧密相关的，不关心科学的预测者就如同不关心食物的厨师一样。科学的独特之处在于，我们关注的是客观世界，这也是使预测变得科学的原因。如果我们只关注方法、准则或是模型，那预测最终必败无疑。

气温变化趋势的真相

如果说阿姆斯特朗的批判没有道理，那么他与戈尔打的赌又该怎么解释呢？阿姆斯特朗的“无变化”预测不仅没有失败，反而还大获全胜了。自从阿姆斯特朗在 2007 年打了那个赌之后，虽然每个月的气温变化幅度极大，但变化却并不稳定，比方说，2011 年的气温就略低于

2007 年的气温。

这种趋势持续了 4 年多的时间。更重要的是，有个难以忽视的真相，2001~2011 年这 10 年间，全球气温根本就没有上升（图 12–9），反而下降了，虽然降幅不那么明显，但实际情况确实如此。

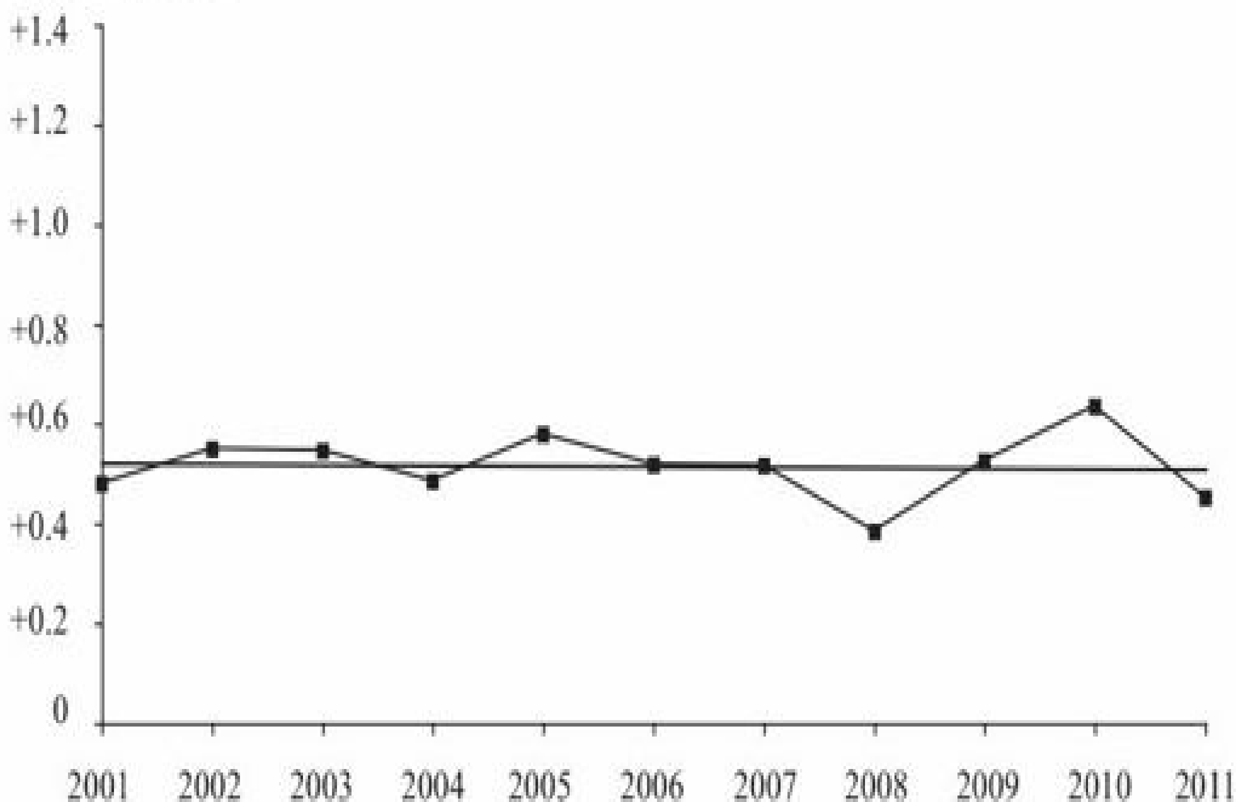


图 **12-9 2001-2011** 年的全球平均气温

这类预测框架有时可能是存心欺骗。比如，如果你将 ENSO 循环期中气温很高的 1998 年作为起点，就很容易得到降温的“趋势”。相反，2008~2018 年这 10 年的气温“走势”就呈现为不断升高，因为 2008 年是相对凉爽的一年。

然而，全球变暖的进程并不是匀速推进的。气温上升的历史是一个明显的长期增长过程，但期间也有保持不变或者负增长的趋势。比如，尽管二氧化碳的浓度一直在增长，除了 2001~2011 年这 10 年之外，1894~1913 年、1937~1956 年、1966~1977 年（图 12–10）这几个时间段几乎都没有升温的迹象。这一问题与金融分析面对的问题有些相似之处：在很长一段时间里股市持续攀升，但这种走势并不能告知你明天、

下周或是明年的股市表现如何。

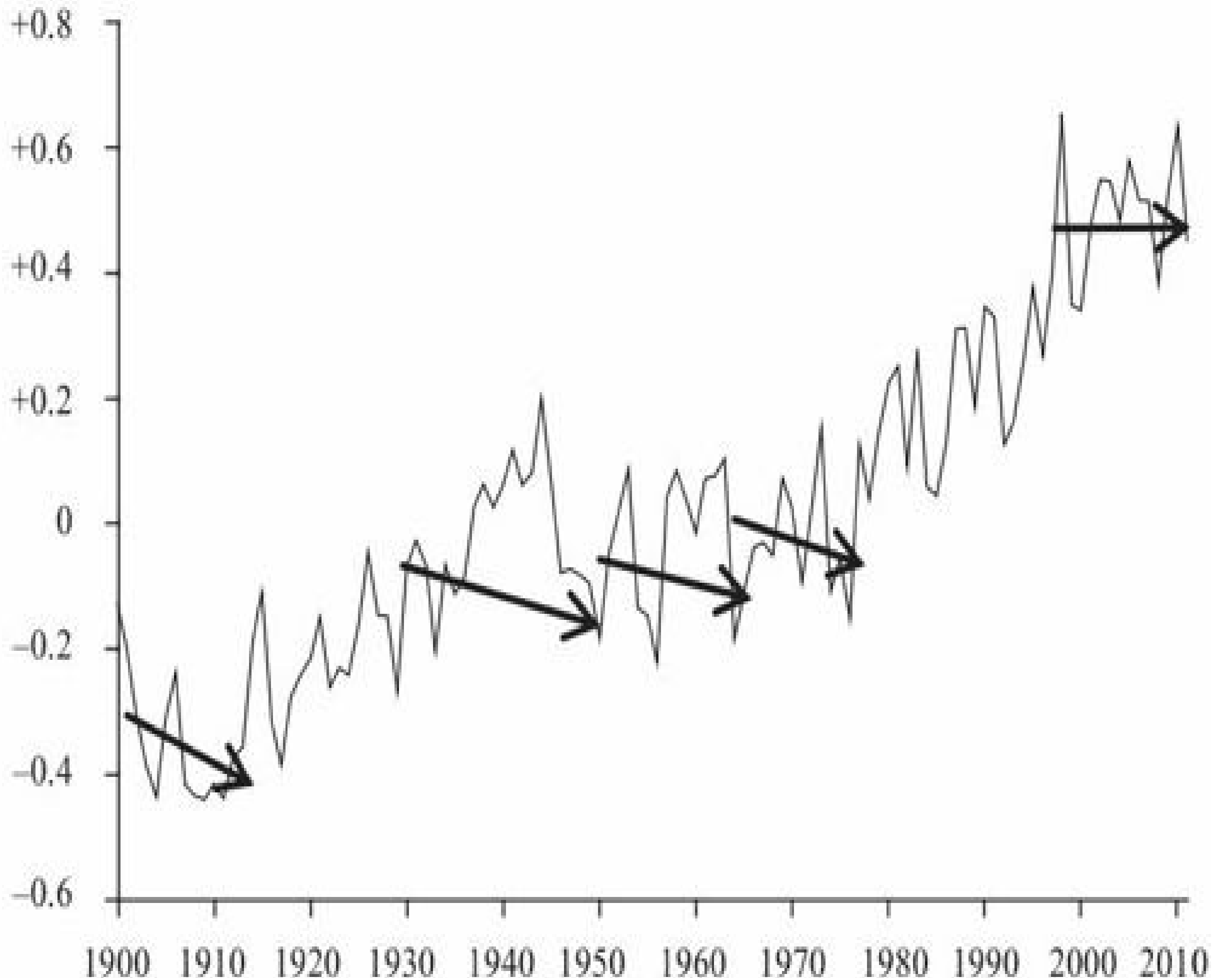


图 **12-10 1900-2011** 的全球气温曾出现短期气温持平和显著下降的趋势

我们可以直接从科学的角度对最近出现的气温保持相对稳定的现象做出解释，比方说，发展中国家不断增长的硫排放量就抑制了气温走高的势头。2001~2011 年的气温虽然没有上升，但比之前的任何一个 10 年的气温都要高。

尽管如此，本书还是鼓励读者认真思考信号与噪声，并找出用百分比或概率表达的预测，这些预测可以真实地体现出我们预测能力的局限性。当预测者对某一个复杂现象的预测表现出十足的信心时，那可能说明他并没有认真思考这个问题，只是将各种数据过度拟合到自己的统计模型当中，或者他并不想找到真相，只是想炒作自己的人气罢了。

阿姆斯特朗和施密特都不愿意对气温走向作对冲预测（即两面下

注）。阿姆斯特朗对我说："我们模拟了 1850~2007 年的气温走向，看到前 100 年的气温变化时，我几乎可以确定与戈尔打的那个赌，我赢定了。"而施密特则认为气温将会继续上升，他还愿意提供诱人的赔率与任何人打赌。"我敢打赌下一个 10 年的气温将会比这个 10 年的气温还高，以多大的赔率打赌都行，即使是 100∶1 的赔率，我也同意。"

前文中提到的统计预测方法，就可以解决阿姆斯特朗和戈尔之间的争论——他们两个的预测都不是十分准确。如果你以 10 年为期限测量气温的发展趋势，那么自 1900 年以来，呈升温趋势的时间占到 75%，而呈降温趋势的时间只占 25%。随着大气中二氧化碳浓度的增加速度不断加快，温室效应的信号也不断增强，气温持平期或下降期的出现频率也会降低。尽管如此，此类情况出现的概率总是存在的，就如 100∶1 的赔率也会存在一样。如果你认为二氧化碳的浓度会以目前（即每年 2PPM）的速度增长，那么根据这一统计方法，在某个特定的 10 年里不会出现净升温的概率为 15%。

不确定性是预测必不可少且不可或缺的一部分。正如我们所知道的那样，有时对不确定性的一个可靠而准确的表达有可能会拯救许多财产和生命。在其他情况下，比方说股票期权交易或者是赌一场 NBA 比赛，你赌的也许就是自己预测不确定性的能力。

之所以需要仔细且明确地对不确定性进行量化，还有另外一个原因，这个原因是科学进步中必不可少的一部分，与贝叶斯定理密不可分。

假设在 2001 年之前，你就已经十分笃信二氧化碳排放会导致气温持续上升。（在我看来，这是恰当合理的，因为我们对温室效应的理论研究和实验证据都支持这样的观点。）你认为全球变暖假说有 95%的概率是正确的。

但是，之后你观察到一些新的证据：在接下来的 10 年里，也就是从 2001~2011 年，全球气温非但没有上升，反而下降了，尽管降幅很小，但的确是下降了。根据贝叶斯定理，此时你应该将全球变暖假说成立的概率预测值下调，而问题在于气温究竟下降了多少。

如果你已经正确地预测到短期气温变化模式中的不确定性，那么对自己估值的下调幅度就不会非常大，反之调整幅度会很大。正如我们所看到的那样，在 10 年内，即使出于气候多样性的原因，全球变暖假说得到了印证，但不出现净升温的概率也会有 15%。相反的，如果气温变化完全是随机且无法预测的，那么就有 50%的概率会出现一个气温下降的 10 年，因为气温上升和下降的概率是一样的。根据贝叶斯定理（表 12–2），无净升温的 10 年会使你很无奈地将自己对全球变暖假说成立

的概率预测值从95%下调到85%。

12–2 贝叶斯定理——全球变暖假说

先验概率		
全球气温上升的初始预测	x	95%
新证据出现：10年里未出现净升温		
如果全球变暖假说是正确的，10年里不出现净升温的概率	y	15%
如果全球变暖假说是错误的，10年里不出现净升温的概率	z	50%
后验概率		
在10年里不出现净升温的情况下，全球变暖假说成立的概率的重新预测	$\frac{xy}{xy+z（1-x）}$	85%

另外，即使你断言10年内气温不会上升的概率只有1%，现在看来你的理论也已经陷入非常糟糕的境地了，因为你的断言过于明确。此时根据贝叶斯定理，你赋予全球变暖假说成立的概率已经下降到28%了。

当我们提出更加自信的主张而又未能实现时，这就为驳斥我们的假说提供了更加有力的证据。一旦出现这样的情况，我们不能责备任何不再信任我们的预测的人，他们不过是根据贝叶斯定理的逻辑做出了正确的推断。

那么，我们为何一开始就要提出更加自信的主张呢？尤其是在这些主张还没有真正被统计数据证实的情况下就这样做，这是为什么呢？其实，人们这样做的原因有很多。在气候争论中，可能是因为这些更自信的主张更有说服力，但这些主张只有在正确的情况下才可能更有说服力。除了被称为气温上升的全球变暖现象，把所有异常天气都认定为人为造成的气候变化，实际是一个筹码很高的赌注，因为这涉及更多政治因素，而不仅是一种科学现象。气候变化会自动显露出来，而不是通过气温上升或海平面上升这些现象体现出来的，对于这一点，人们很难达成共识。显然，把每一次降雪都认作驳斥气温上升的证据是十分荒唐可笑的。

政治与科学的针锋相对

气候学家面临的两难境地是，全球变暖是一个需要在短期内解决的长期难题。这是因为二氧化碳在大气中长期存在，而我们今天所作的决

定将会影响到子孙后代的生存问题。

若是在一个完全理性且充满仁爱的世界里，这样的问题就不会那么令人烦恼。但是，我们的政治机构和文化机构并不能很好地处理这样的问题。比如，美国国会在面对 4 年一次的美国总统大选或是美国经济为满足每季度的赢利预测而承受巨大压力时，都有些力不从心。气候学家以各种方式应对这一挑战，有些卷入了政治争论，而有些对此保持距离。宾夕法尼亚州立大学地球系统科学研究中心的负责人迈克尔·曼曾经是论战的中心人物。在“气候门”事件中，他为英国国家气象局提供的气温记录被黑客入侵，这一组织隶属于东安格利亚大学的气候研究小组。一些怀疑论者断言，迈克尔·曼和其他科学家密谋操纵了气候研究小组的气温记录。

相关事实是，科学家同行为他们洗清了嫌疑，气候研究小组的气温记录同其他科学家的记录基本一致，但是在黑客入侵的邮件中，迈克尔·曼和其他科学家明确表达了他们对科学理解过程中的公共关系因素的关注。在一个秋高气爽的下午，我到他的大学办公室里拜访了他，我们谈了约两个小时。

迈克尔·曼对全球变暖背后的科学道理有其独到的见解。与大多数气候学家一样，他并不怀疑气候变化的理论机制，但他对气候模型做出的预测却持怀疑态度。

迈克尔·曼对我说：“对科学任何可靠的评价我都会认同。有些事情我们非常了解或是有几分了解，但对于有些事情我们充满了不确定，甚至对其一无所知。”

“在我看来，这次居心叵测的公开对话带来了一系列不幸后果，其中之一就是，我们正在把时间浪费在争论某一个科学家已经广为接受的观点上，而我们本可以对切实存在的不确定性进行开诚布公的讨论。”

与施密特同在 RealClimate.org 网站上发表博文的迈克尔·曼发现，自己已经卷入了与哈兰学会这样的组织的堑壕战中。他引用了《自然》杂志一篇社论中的一句话：“我们与这些人在街头对峙”。街头对峙的长期目标，是让公众和决策者认识到采取行动抵抗气候变化的紧迫性。过度自信的预测者总是误把自己的信心当作真实情况，如果全社会都习惯了他们的这种做法，那无论是公众还是预测者，都不会再把不确定性看作制胜策略了。

迈克尔·曼对我说：“当你对不确定性的位置相当明确时，就要划定界限，不要让我们的言语中充满不确定性，以致没有人会认真倾听。作为一个团体，不将自己的观点表达出来是不负责任的。总有一些人乐于填补空缺，但他们用来填补空缺的却是不真实的信息。”

实际上，迈克尔·曼口中的“街头对峙”主要是在 Real–Climate.org 网站这类“共识”网站和 Watts Up With That.com 网站这类“怀疑”网站之间展开，围绕每天的最新杂志文章、天气模式或政治论战进行。两家网站各持己见，谁都不愿让步。一朝是喷气式飞机，永远都是喷气式飞机，死不让步。

我并不是说两个网站平分秋色。在全球变暖这一科学争论上，真理似乎只站在一边：温室效应确实存在，由于人类排放二氧化碳气体，温室效应还会继续加剧。这很可能会使地球升温。温室效应的影响并不确定，但一定不会带来什么好结果。

然而，如果只有几个人信服科学，那么“街头对峙”似乎就认定了我们即将解决的是政治问题。实际上，我们还需要很久的时间才能让人们信服科学。在哥本哈根气候大会上，理查德·鲁德对我说：“我曾经得出一个结论，人们必须想出低碳的方法。”但他预感到，193 个成员国几乎不可能找到彼此都接受的方法。

与此同时，在过去几年里，美国公众对全球变暖正在发生这一事实的信任度已经下降了。即使对气候变化带来的影响完全认同，但一些州和国家减少碳排放的计划做得就没有其他地区或国家好。华盛顿州州长克里斯蒂娜·格雷瓜尔对我说：“在产煤州有一些非常进步的民主党州长，你说他们对此是否紧张呢？”

这些问题并非气候争论所独有，而我实在不知道该如何解决这些问题。但是我知道，科学与政治之间存在着根本性的差异。实际上，我已经越来越肯定这两者是彼此对立的。在科学中，进步是可能的。如果我们相信贝叶斯定理，一旦做出预测、观念得到检验和改良，科学进步便是必然的。科学进步的道路并非一帆风顺，有些深受好评（甚至“达成共识”）的理论后来也被证明是错误的，但不管怎样，科学总是向着真相不断前进的。

与科学不同，在政治上我们似乎越来越难达成共识。自罗斯福“新政”实施到 20 世纪 70 年代以来，美国众议院两党之间的分歧已经有所减少，但在 2011 年，分歧又开始加大，在过去至少 100 年的时间里，两党之间的分歧达到了顶峰。共和党已经远离了共识，而民主党也在一定程度上有所偏离。

在科学领域，很少看到所有数据都集中到一个明确的结论上的情况。真正的数据非常嘈杂，即使理论是无懈可击的，信号有时也会有所偏离，更何况根据贝叶斯定理，无懈可击的理论是不存在的。科学是一个渐进的过程，需要进一步提炼与检验，这是科学怀疑论的全部。

而在政治领域，人们从不让步。某人言语不当会被视为表现失礼，

确实如此。人们期待政党对于经济、社会及没有内在关系的对外政策等问题保持平等的信念。但是随着世界变得越来越接近，民主党和共和党的纲领也会变得越来越不完善。

正是因为政治与科学之间的争论还会持续几十年，气候学家最好还是从街头对峙中撤出，避免跨越卢比孔河，从科学界踏入政坛。在科学领域，可疑的预测更有可能会现出原形，真相也更有可能占上风。在政治领域，真相并不能享有特权，它可以是任何人的猜测。

美国政治系统失调的状态是解释我们对美国未来前景持悲观态度的原因，而科学技术的不断发展则是解释对美国未来前景持乐观态度的原因。美国是一个善于创造的民族，它的专利数量数不胜数，还拥有世界上最好的大学和研究机构，不论是医药还是信息技术，美国的公司都能占据市场的主导地位。如果让我在思想的锦标赛和政治的牢笼赛之间做出选择，我知道我会选择哪一个，在我自认为预测正确的时候尤其如此。

第十三章
恐怖主义：比“9·11”更严重的恐怖袭击事件会发生吗？

富兰克林· 德拉诺·罗斯福称 12 月 7 日是美国“活在耻辱中的一天”。1941 年 12 月 7 日发生的“珍珠港事件”是一个多世纪以来美国本土第一次遭受他国攻击，这次事件对美国人内心的冲击不亚于 60 年后世贸中心大楼被袭，这使得美国人意识到，曾在地球另一边的潜在敌人已变成实实在在的威胁。更令人讶异的是，事后证实珍珠港事件似乎是可预测的。

事发前，有许多信号都显示珍珠港可能将遭遇偷袭，甚至已兵临城下。1941 年 11~12 月，美日关系急剧恶化。日本扩张领土的野心膨胀，美国总统罗斯福将太平洋舰队从圣迭戈移防至珍珠港，以威慑日本，而这步棋却使美国太平洋舰队成为该区域的敏感目标。与此同时，日本海军频频更换无线电代码，并从中国和东南亚沿海调派大批军队和战舰驶向珍珠港，火药味愈来愈浓。

但最不祥的信号还是暴风雨前的平静。虽然美国情报官员已成功破译日本密电码“紫色”（PURPLE），可解读日方 97%的外交密电，但是美国破解日本军事密电的尝试却鲜有成功。尽管无法解读这些军事密电的内容，但美国能侦测并追踪到密电发出的位置。日军航母舰队出海时不断发送的电报实际上已经暴露了它的行踪。

然而，从 1941 年 11 月中旬开始，日军无线电全部静默，致使美国对日军航母的行踪一无所知。20 世纪 40 年代还没有全球侦测卫星，只有落后的雷达。在广袤的太平洋上空执行侦察任务，成本又十分高昂，因此只能在基地附近方圆 483~644 公里内展开无规律的巡逻。日方的舰队由数艘战舰组成，每艘战舰都有 6 个足球场那么大。而美军当时最好的侦察手段就是无线电信号，如果没有这些信号，他们就无法获得日军整支舰队的踪迹。

当时美方情报机构中有许多人认为，日军航母彼时距美国领海很近，在这种情况下美军可以依靠不同的通信工具交替监控日舰行踪。当然还有另一种可能性，即日军舰队已经冒险闯入太平洋海域，避开了美国海军的侦查设备。

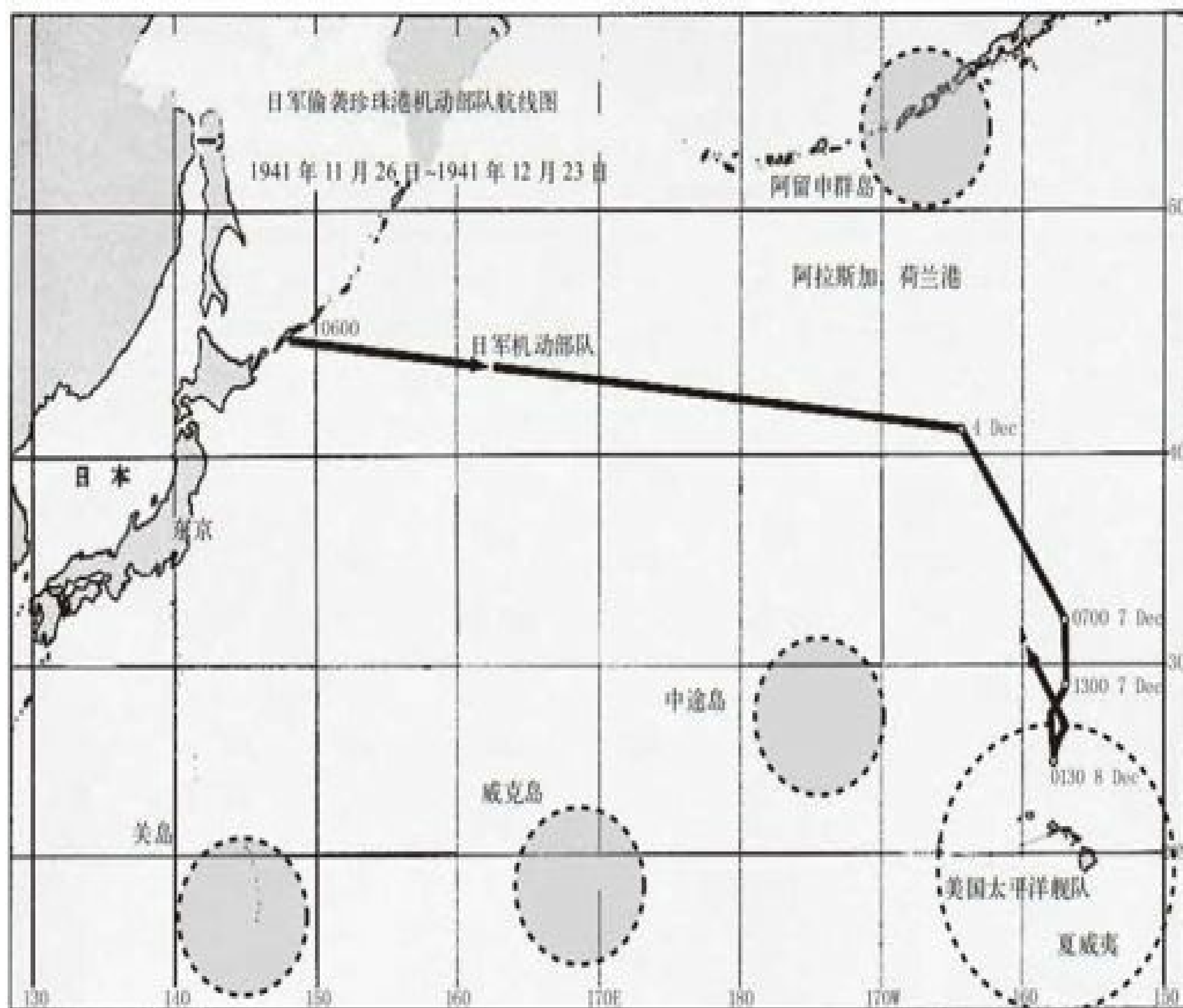
其实，日本航母舰队当时正在开往夏威夷的途中。日军早已制定好精确的航线，就好比在美式足球比赛中，一个踢四分卫的球员发现了对

方防卫中的漏洞，便借此发起进攻一样，当时的日军舰队正在穿越美军防线的盲点。日本舰队最初沿东南方向行驶，几乎是不偏不倚地行驶在中途岛和阿拉斯加荷兰港两座海军基地的中轴线上。1941 年 12 月 4 日，日军航母舰队抵达西经 165 度的位置时，突然 45 度转弯，取道向南，直奔夏威夷。3 天后的清晨，日军对珍珠港发动了袭击，这次袭击使美军损失了近 2 400 名士兵，4 艘美军海军战舰被击沉。

1941 年 12 月 8 日，美国国会以 470∶1 表决通过对日宣战，正式加入第二次世界大战。

信号无处不在

2001 年 9 月 11 日，美国航空 77 号班机被劫持并撞向五角大楼西翼，机上 59 名乘客全部罹难。这次袭击也夺去了美国前国防部部长唐纳德·拉姆斯菲尔德身边 125 名同事的性命，这件事让拉姆斯菲尔德想到了珍珠港事件。1941 年 12 月 7 日，当时只有 8 岁的拉姆斯菲尔德正在用广播收听他钟爱的芝加哥熊队的比赛，突然插播的日军偷袭珍珠港的新闻中断了那天的赛事播出。



注：以下时间均采用日本东京时间

4 Dec：12月4日

7 Dec：12月7日

8 Dec：12月8日

图 **13-1** 日本舰队偷袭珍珠港路线图

2012 年 3 月，我前往拉姆斯菲尔德位于华盛顿的办公室采访他，去之前已经有人提醒过我此次采访未必会达成所愿，所以不要抱过高期望。拉姆斯菲尔德身高 1.73 米，当时已年近 80 岁，体格算不得多么壮硕，但在待人接物和言谈举止等方面都颇有震慑力。拉姆斯菲尔德出生在伊利诺伊州埃文伊顿市，在那里就读公立高中，后毕业于普林斯顿大学。大学时，拉姆斯菲尔德曾是摔跤队的主力，毕业论文的选题锁定了美国总统的权力。大学毕业后，拉姆斯菲尔德先是在美国海军服役，后当选为美国国会议员。在拉姆斯菲尔德办公室的墙上挂有他曾任职过的 4 个部门的徽章和纪念品：他是美国历史上第一位两次担任国防部部长

的人，第一次是在 1975~1977 年福特总统任内，第二次是 25 年后在小布什总统任内。

此次采访之前，我就了解到拉姆斯菲尔德对珍珠港事件颇感兴趣，于是将本书的详细大纲交给了他的参谋长——年轻的基斯·乌尔巴汗。幸运的是，采访当天拉姆斯尔德的心情不错，他已经仔细看过我提交的那份大纲。一见面，他便将一本著作的前言影印件递给我，这份前言摘自罗伯塔·霍斯塔尔 1962 年出版的《珍珠港：征候与决心》一书，这本书概述了日军的偷袭让美国军队和情报官员如此震惊的各种原因，认为比毫无准备更坏的情况是，我们把无知当成博闻，结果变得更加不堪一击。

拉姆斯菲尔德说："事后发现，驻扎在珍珠港的美军事先准备好要应对的事，结果都没有发生。例如，夏威夷有许多日裔，美军为防止他们搞破坏把所有飞机密集地并排停靠在一起，认为这样就可以更好地保护飞机。但是当轰炸机来袭时，如此停放会使飞机更易遭受攻击和被摧毁。"

拉姆斯菲尔德指出，在"珍珠港事件"前，美军普遍认为美国本土的日裔最有可能对美军飞机和船只发动袭击。这种忧虑笼罩着整个夏威夷群岛。美军认为岛上的 8 万名日裔不仅会偷袭军事基地和无线电台，还会破坏菠萝园和乳牛场。在这种情况下，不管是否符合逻辑，任何信号都会朝这个方向解读，美方一直在提防日裔耍花招，于是，把飞机和舰船紧密地停在一起，以利于监控。

同时，美方推测，如果日本正在策划攻击，就得先攻占苏联（东北部地区），或英属亚洲殖民地，而苏联和英国早已卷入第二次世界大战，日本已经树敌众多，为什么还要激怒美国这只沉睡的雄狮呢？美国人并没有发觉日本人一定会把自己拖进战争的旋涡，其实这些日本人想趁美军最疏于防备的时候发动攻击，重创美国海军。当时的日本政府不愿放弃任何领土扩张的希望，而美军却没能从日军的行动中看出任何战争的苗头。

在霍斯塔尔看来，所谓信号就是能够向我们传递敌人实际意图的有用证据；而本书则认为，信号是统计或预测问题背后真相的表征。霍斯塔尔关于"噪声"的定义与我的略有不同。霍斯塔尔所说的噪声指的是相互矛盾的信号制造的声音，而我则倾向于用这个词指代那些易被误认作信号的随机模型。在情报分析领域，信号缺失意味着将有重大事件发生（例如第二次世界大战时，日本航母舰队的无线电静默暗示着它们意欲开赴夏威夷，偷袭珍珠港），而信号过多又会使意义识别工作异常困难，这些信号可能会被淹没在震耳欲聋的噪声中。

下面一组图分别由 10 个信号组成，每个信号都由一个简单的数学函数得出，即正弦波。在图 13–2A 中，有一个信号特别突出，与其他信号不同。这张图呈现的就是我们遭遇一次袭击或其他失败的预测后的情形，我们所看到的世界就是这个模样。我们会看到这个信号在书面记录中有，在模型里有，在各种先兆里有，信号无处不在。“珍珠港事件”和“9·11”恐怖袭击事件之后都有少数美国人声称，此类事件的模型是如此清晰明确，政府事前肯定已经得知要发生袭击，因此他们推断政府肯定参与策划或实施了这些袭击。

但在事发前，我们所看到的模型通常不是图 13–2A 所展示的那样，而更像图 13–2B 所展示的，如同一团杂乱无章的意大利面。正如霍斯塔尔所写的那样：

> 事后更容易从一堆无关的信号中捋出有用的信号。当然，事发后的信号总是相当清晰，因为灾难已发生，我们便可以看出该信号究竟预示着怎样的灾难。但是在事发前，信号总是模糊不清，让人难以琢磨，此时观察者身处“噪声”包围的环境中，也就是说，某场特殊灾难的预测过程总会伴有各类无用和无关的信息。

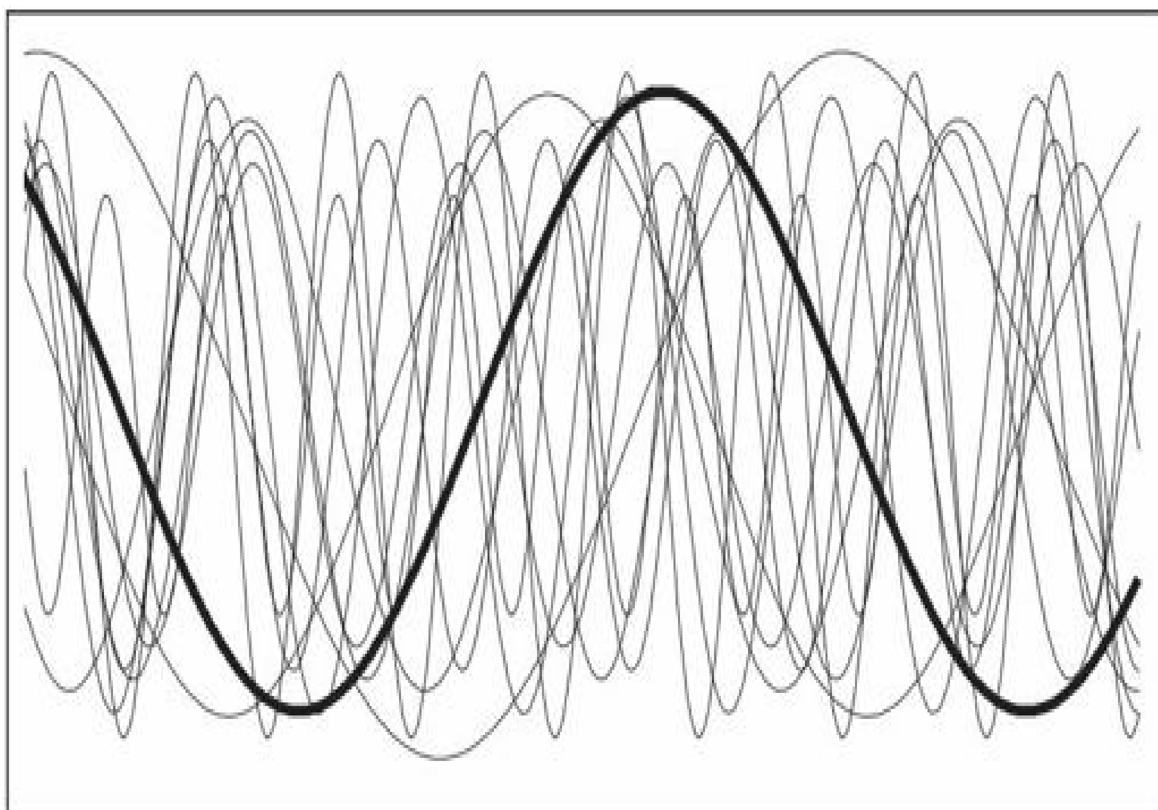


图 **13-2A** 一个强信号和其他弱信号

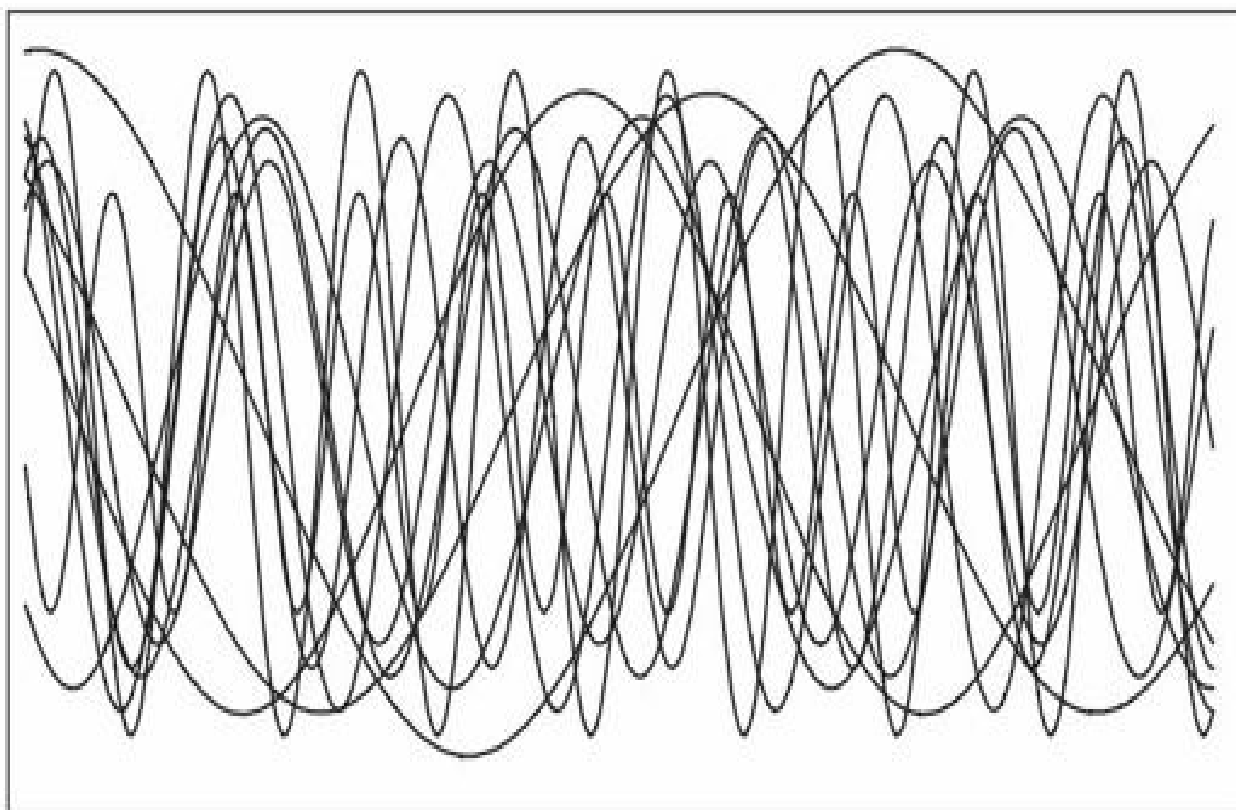


图 **13-2** 图**B** 相互矛盾的无强弱差别的信号

在这样的情况下，重要的就不是我们的信号探测能力了。假设我们已经具备了基本的信号探测能力，并在“珍珠港事件”或“9·11”恐怖袭击事件这样的重大事件发生前发觉大量信号，相关信号将被存入档案柜或输入电脑数据库，但大量不相关的信号也会存入其中。我们真正需要的是，从杂乱无章的信号中分离出有用信息的分析能力。

哪些信号更重要、更值得关注，我们通常对此都有自己的见解，持有自己的见解在一定程度上是很有益处也很有必要的。我在前文中已经详细地指出了在分析数据时因为忽略环境因素而引发的各种问题。我们没有做出什么有用的预测，却总是在各种预测模型上举棋不定、跳来跳去。

然而，我们关注的环境因素本身也会有失偏颇，为了一己私利。正如莎士比亚借西塞罗之口警告尤利西斯·恺撒一样，“人们照着自己的意思解释一切，实际上却与事物本身的目的完全相反”。有些信号会推动我们推崇的世界理论向前发展，或者预示着更加乐观的结果，对于这些信号我们也许会十分关注。或者我们可能只会关注那些符合官方口径的信号，就像“珍珠港事件”中美军更愿意相信珍珠港面临的威胁来自美国本土的日本人，而不是日军的空袭。

你不知道的不代表不会发生

拉姆斯菲尔德最喜欢霍斯塔尔那本《珍珠港：征候与决心》的前言，这个前言是诺贝尔经济学奖获得者托马斯·谢林写的，正是谢林把约翰·纳什的博弈理论应用到美国国家安全领域的。谢林写道，人们常常误以为“不熟悉”就是“不可能”：

> 人们在作规划时，总会将不熟悉和不可能相混淆。那些我们从未认真考虑过的偶发事件看起来总是很奇怪，那些看起来很奇怪的事情则常被视为不可能发生，而不可能的事情就不需要认真考虑了。

美国和欧亚大陆远隔重洋，门罗主义颁布以来，美国与美洲其他国家的关系相对良好，一直以来，美国都极少遭受外国袭击。因此，“9·11”恐怖袭击事件、古巴导弹危机这样的特例对美国人而言属于很大的冲击。“珍珠港事件”之前，美国领土最后一次受到外国袭击是发生在美国第二次独立战争期间。欧洲和亚洲人民经历过多次战争的摧残，但美国不同，美国人民过去没有在战争废墟中生活过。

但是，夏威夷的情况比较特殊。这座城市位于太平洋中部，檀香山距离东京（6212 公里）比距离华盛顿（7 765 公里）还要近。因其特殊的地理位置，外加美国海军舰队的出现，夏威夷成为日军的重点打击目标。美国对外国军队袭击美国本土的情况不熟悉，因此对这次威胁掉以轻心了。

或许我们可以依据以下线索进行逻辑推理：

1. 美国几乎没有遭受过袭击。
2. 夏威夷是美国领土的一部分。
3. 因此，夏威夷也不可能遭遇袭击。

这是一种非常错误的思维方式。正如前文中所讲，无样本依据的冒险预测常常会失败。美国很少遭受袭击，这一事实是一个经验观察，而不是什么铁律。比方说，内布拉斯加州从未遭受外国的暴力袭击，这一事实并不能作为铁证运用到夏威夷身上，因为夏威夷位于大洋中部的偏远位置，周遭的战争局势又给它带来了安全隐患。

但是，至少这类错误的思维方式包含了一些“思考”。如果我们曾经仔细检查过这一思考过程，也许就会意识到那些假定条件是多么宽泛、

多么不严谨。谢林建议我们应该提问时问得更深入一些。当我们对某种可能性不是很熟悉时，连想都想不到它，我们会对这种可能性形成意盲，医学上称为病觉缺失征。这种疾病的部分生理反应阻碍病患意识到生理患病的事实，许多老年痴呆症患者常呈现出这种状态。

当预测出现这一病症时，我们就需要与自己的本能做最激烈的抗争，勇于承认我们有哪些无知之处。

“9·11”恐怖袭击事件是“已知的未知”？

> 有些事是已知的已知，即我们清楚自己知道它。有些事是已知的未知，即我们清楚自己不知道它。有些事是未知的未知，即我们并不清楚自己不知道它。
>
> 唐纳德·拉姆斯菲尔德

拉姆斯菲尔德这番著名的“未知的未知”言论发表于 2002 年，是他在“伊拉克存在大规模杀伤性武器问题”的新闻发布会上说的，这一言论是谢林所关心的“将不熟悉当作不可能”这一问题的延伸。如果我们能提出一个疑问并给出标准答案，这就是已知的已知。若自问一题却无法给出准确答案，这就是已知的未知。而未知的未知则是我们连想都没想到应该提什么问题。“未知的未知是我们知识的鸿沟，但我们却不知道它的存在”，2011 年，拉姆斯菲尔德在他的回忆录中这样写道。

有时，“未知的未知”这一观点会被人误解。这一术语常被用来预测那些相当具体（但也相当难以预测）的威胁：

> 赌尼日利亚在不久的将来会出现危机，这是一个有把握的赌局——这样一个未知的未知对美国以及全球安全（重点强调全球安全）都是非常重大的预测。

预测尼日利亚会发生恐怖袭击是相当准确的（上面这段话写于 2006 年， 3 年后尼日利亚人乌玛·法鲁克·阿卜杜牧塔拉布在荷兰阿姆斯特丹机场进入飞往底特律的航班，他将炸药藏在自己的内裤里携带上飞机，并伺机引爆）。然而，这段话存在语义错误。只要你还能列举出危险因素或不可预测的因素，那就是在表达已知的未知。能说出你所不知道的事情，就是进步的标志。

但正如我们发现的那样，几乎没有什么事情能简单地划分为可预测和不可预测这样的二元类别。即使你不知如何十分肯定地预测某事，但

也许你能够对威胁进行估算或预测。估算可以精确也可以粗糙，预测可以准确也可以失准，可以聪明也可以愚钝。但至少你对问题有所警觉，这样通常就能有所收获，例如，我们不知道尼日利亚恐怖分子对美国发动袭击的可能性究竟有多大，但我们知道该国对美国的威胁很可能比卢森堡要大。

我们对世界有限的认知让我们备受挫折，当我们根本无法预测时，麻烦就出现了。“未知的未知”是指我们从未考虑到的偶发事件。面对这种情况，我们会有些思维障碍，或者说我们的经验不足以想象到此类事件的发生，于是，便认为这件事情根本不可能发生。

在评估接收到的恐怖分子的信号时，这种观念给我们带来了巨大的风险。如“珍珠港事件”发生之前的情况一样，“9·11”恐怖袭击事件发生之前也有许多信号：

·已有诸多警告说飞机可能会被用作武器，这些警告包括 1994 年阿尔及利亚恐怖分子企图劫持飞机冲撞埃菲尔铁塔，1998 年一个与“基地”组织有关的团伙企图利用载有杀伤性炸药的飞机撞击世贸中心大楼。

·世贸中心大楼之前就曾是恐怖分子袭击的目标。1993 年，曾在阿富汗“基地”组织受训的马泽·尤瑟夫及其同伙制造的一场爆炸造成 6 人死亡，当时他们意在炸毁双子塔。

·众所周知，“基地”组织是一个极其危险、手段多变的恐怖组织。该组织实施了多次大规模恐怖袭击，包括：1998 年炸毁美国驻肯尼亚和坦桑尼亚大使馆，造成 224 人死亡；2000 年，又袭击了美国军舰“科勒”号。

·2001 年 7 月，美国国务卿康多莉扎·赖斯曾收到过关于“基地”组织频繁活动的警报，该组织已经将袭击目标从其他国家转移到美国。“我的第六感告诉我这些，我能感觉到要出事了，而且还是一件大事。”中央情报局局长乔治·特尼特见到情报后如是说道。

·一位名叫扎卡里亚斯·穆萨维的宗教极端主义者在 2001 年 8 月 16 日——恐怖袭击发生前不到一个月的时候——在美国被捕，来自明尼苏达州的一名飞机培训师向美国政府举报他，认为他的形迹可疑。穆萨维的飞行训练时长还不足 50 个小时，并且从未单独飞行过，却试图参加波音 747 客机的模拟训练，一个还要很久才能获

得飞行证的学员提出这样的要求甚是诡异。

事后再辨别这些重要的信号，会容易得多。美国国家安全部门必须从数万条甚至数十万条潜在的警告信息中披沙拣金地找出有价值的信息，而其中大多数信息都没有什么价值。

另外，“9·11”恐怖袭击事件的谋划是相当周密的，恐怖分子并没有受到什么阻挠就轻松得逞了。19 名恐怖分子“成功”混入航空运输系统，“成功”地劫持了 4 架飞机。其中 3 架又“成功”袭击了目标，只有 93 号航班没有被恐怖分子利用，这完全是因为乘坐该飞机的乘客在得知飞机被劫持后英勇搏斗，成功护卫住飞机驾驶舱，最终飞机在宾夕法尼亚州的尚克斯维尔附近坠毁。我们不仅当时没有觉察到这次密谋袭击，即使到了现在也没有彻底弄清楚事情的原委。

“9·11”恐怖袭击事件调查委员会的报告中总结出美国存在的 4 类体制问题，就是这些问题使得政府无法领会这些信号的重要性。这 4 类问题包括政策问题、能力问题、管理问题，还有最重要的一类问题，就是想象力问题。这些信号与我们熟知的恐怖分子的行为方式的相关假设不相吻合，所以我们总是把它们当成耳边风，左耳进、右耳出，而没有真正将其记录在案。

北美航空航天防御司令部曾经提议进行一次有关“被劫客机袭击五角大楼”的军事演习。但是，这一想法因为“太不现实”，最终未被考虑。而且当时人们认为，即使这种不太可能发生的事件真的发生了，遭劫的飞机也应该来自海外航空港，而不是美国本土的机场。（颇具讽刺意味的是，这个设想和“珍珠港事件”时美国所犯的错误恰好相反，当时美军否定来自国外军事力量袭击的可能性，而且当时防御政策的制定者都在担心本土暴乱。）

同样，美国人也很难想象会遭到自杀式爆炸袭击。美国联邦航空管理局的政策称，劫持飞机会导致局势紧张，也会使一些国际航班绕过中东地区飞行。但是，该政策并不认为恐怖分子真的想要摧毁飞机或是杀死乘客，只是以此作为一种谈判策略。因此飞机的驾驶舱并不是完全密闭的，而且实际飞行时通常不上锁。

但众所周知，自杀式袭击由来已久，当然包括日本在第二次世界大战期间“神风”特攻队的飞行员。另外，就在“9·11”恐怖袭击事件发生前的几年里，自杀式袭击越来越普遍。据一个恐怖事件的相关数据库记录，20 世纪 80 年代总共发生此类袭击 31 次，但是进入 21 世纪之后，自杀式袭击频发，仅 2000 年一年就发生了 39 起自杀式袭击，其中包括发生在也门的对美国军舰“科勒”号的自杀式爆炸袭击。

然而，第二次世界大战已是遥远的记忆，而且当时大多数自杀式袭

击都发生在中东地区或是第三世界国家。被丹尼尔·卡尼曼称为“可得性法则”的心理捷径——我们常常会高估那些在时空上离我们更近的事件的发生概率，而那些离我们较远的事件，我们常会低估其发生的概率——很可能会扰乱我们的判断。

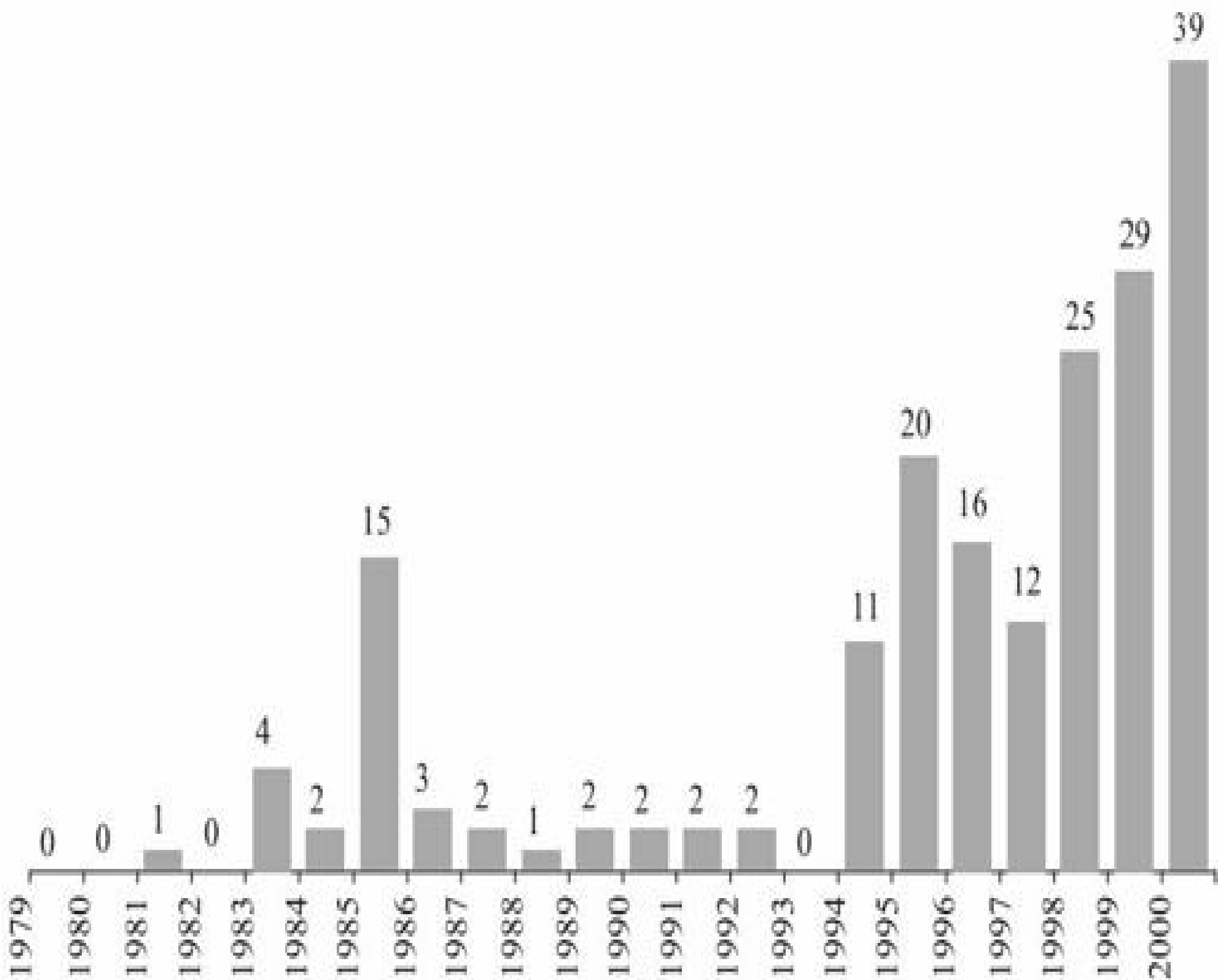


图 **13-3 1979-2000** 年自杀式袭击数量

“如果人们不愿去死，那你就可以按照常理对其行为进行预测，”拉姆斯菲尔德对我说，“但是，如果人们视死如归，或者觉得死亡是一种殊荣，觉得死得其所，那他们就不会按常理出牌了。”

“9·11”恐怖袭击事件的阴谋并非经过评估被认定为不可能而遭到否决的假设，实际上，从一开始我们压根儿就没有想到会发生这样的事，这样的阴谋对我们来说太陌生了。拉姆斯菲尔德在他的回忆录中提到：“9·11”恐怖袭击事件是一个未知的未知。

拉姆斯菲尔德的参谋长乌尔巴汗对我说：“中央情报局会告诉你，‘9·11’恐怖袭击事件这样的阴谋对他们而言并不是完全陌生的，但

很多细节是在袭击发生后被披露出来的，属于典型的后见之明。”

他补充道：“我的意思是，‘基地’组织可以说是一个已知的未知。此外，前中央情报局局长特尼特在其著作中曾经提到‘9·11’恐怖袭击事件的级别，称‘9·11’恐怖袭击事件与其他任何袭击都大不相同。这次袭击相当严重。”

确实，“9·11”恐怖袭击事件的危害巨大——2 977 名无辜的民众被夺去生命——这一点彻底颠覆了我们此前对恐怖主义的认识。甚至那些最为关心“基地”组织活动的人，比如特尼特、国家安全委员会反恐主席理查德·克拉克，他们都没想到“9·11”恐怖袭击事件的影响会如此巨大。比方说，克拉克递交给赖斯的意见书便是恳请她提防可能导致数百名美国人丧生的事件，却没料到美国会发生这样一场致数千人丧生的大事件。

“9·11”恐怖袭击事件发生之前，西方国家遭遇过的最严重的恐怖袭击发生在 1985 年，一伙锡克教极端分子将炸药藏在一架由德黑兰飞往蒙特利尔的印度航空公司的飞机上，造成 329 名乘客死亡，这个数字仅是“9·11”事件死亡人数的 1/10。在“9·11”恐怖袭击事件之前，美国本土遭遇的最严重的恐怖袭击是发生在 1995 年的俄克拉荷马州首府俄克拉荷马城的爆炸案，爆炸针对一座联邦办公大楼——艾尔弗雷德·P·默拉联邦大楼，那次事件共造成 168 人死亡。

但是，“9·11”恐怖袭击事件并不是一个完全无法预料的离群值。尽管这次事件的特殊性无法提前察觉，尽管确实很难预测到这些特殊性，但是，我们本该预测到类似“9·11”恐怖袭击事件这类伤亡巨大的袭击是有可能发生的。

用数学的方法研究恐怖主义

在这里，我们会用抽象的数学方式来看待恐怖主义，这个方式似乎会让人觉得不太舒服。需要明确的是，这种方式并不能代替情报机构的信号分析工作。但这类思维方式能够弥补一些现存的盲点，使我们可以更准确地估计恐怖主义的整体危害。对现有数据进行分析，可以让我们更加清楚地意识到未来的风险。

2008 年，我受邀在一次由战略与国际研究中心举办的会议上发言，该研究中心是一个位于华盛顿特区的外交政策智囊团（即华盛顿智库）。这次会议召开的时机再糟糕不过了——正值 2008 年美国总统大选前两周——但在得知与会成员包括国家安全部门人员之后，我觉得自己应该参会。

此次会议聚集了来自不同领域的专家，希望能借助一系列头脑风暴会议“跳出固有的框框”，为预测及阻止恐怖袭击提供一些创新性见解。

与会成员有可口可乐公司的营销总监、纽约警探、为美国婚恋交友网站设计计算程序的工程师，还有我。（当然与会人员中更有大量专家，他们的工作显然和恐怖主义联系更为紧密，他们效力于美国国会、军队和华盛顿附近的国防契约企业。）

发言时，我简单地介绍了一下自己的职业，讲了自己在棒球和政治领域的预测成果。与会成员听我讲这些情况时还算客气，但随后的提问环节就不那么客气了。有人对我说："纳特，你说的预测手段非常好，但鬼知道我们要怎样把这个手段应用到反恐行动中去呢？"（我对原话稍作了修改。）

不过坦白说，我在会上介绍的方法对国家安全分析而言价值不大。棒球和政治领域数据丰富，能给出令人满意的答案。棒球比赛每年有数千场。美国总统大选发生的频率低一些——在预测时需要更多细心——但是每次美国总统竞选时，都会涌现数百个民调结果。所有这些数据都向公众开放，有些收费低廉，有些则完全免费。

但是，恐怖主义预测领域的情况看似完全不同，我们所关心的"9·11"恐怖袭击之类的事件发生频率相当低。同时，恐怖组织还努力隐瞒他们的计划，"基地"组织在这一方面做得尤为出色。在恐怖主义预测领域，正如"珍珠港事件"发生前的状况一样，有时信号的缺失比出现恐怖袭击的征兆更令人不安。如果中央情报局能够侵入激进组织网上聊天的网站，就会发现他们最初阶段的聊天记录有些随意，因为在这一阶段，像"基地"组织这样的恐怖组织希望能招募一些天真的新面孔。但是，开始密谋袭击或计划升级时，他们的谈话基本上都是在线下进行的。

在微观层面上，即个体恐怖分子的行为或是某个单独的恐怖袭击计划，"魔法子弹"这样的解决方案不大可能存在。相反，情报机构得从一团意大利面似的信号中分类寻找有价值的信息，意大利面的比喻我在前文中已提到，在战略与国际研究中心举办的会议上，我曾和一名专家交谈过，他提出了另一个比喻：探寻恐怖主义的阴谋比大海捞针还难，这个过程也好比是从一大堆针中找出特定的一根。

对于某些问题，用具体问题具体分析的方法很难做出预测，而当我们跳出细节从宏观角度审视时，问题就会变得井然有序一些。在这种情况下，从外部审查恐怖主义的数学特征效果不错。

来自科罗拉多大学的 30 多岁的亚伦·克劳赛特教授拥有物理和计算机的学科背景，曾发表过数篇数学论文，其涵盖范围从鲸类进化到多人角色扮演游戏中的网络动力学研究，内容极其广泛。情报界长期以来一直有偏爱"阿尔法男"的传统，对数学人才情有独钟，而这位 30 多岁

的“神人”克劳塞特的研究兴趣太过宽泛，情报界好像对他颇为讨厌，而人们对他展示的成果也是褒贬不一。

克劳塞斯在电话中对我说：“有人说这种方法就好比一缕新鲜空气，但只有少数人这么认为。大多数人看过之后会说，‘你打算用数学的方法？有点奇怪啊。’”

其实，克劳塞斯的方法相当简单，至少这种方法还有些后见之明的好处。他研究发现，恐怖主义的数学特征和本书讨论的另一个领域——地震的数学特征类似。

假设你生活在地震频发地区，如加利福尼亚州。几十年来，在你所经历的地震中，4 级地震频率稳定、定期发生，5 级地震一年会发生几次，而 6 级地震发生的次数屈指可数。你的房子不能承受 7 级地震，但可以抵御 6 级地震，于是你便称自己没有什么好担心的，这样的结论正确吗？

当然不正确。地震遵循幂律分布，据此定律，发生了 5 级和 6 级地震便预示着有可能发生更大级别的地震。如果时间跨度够大，更大级别的地震其实是不可避免的。大地震终会爆发，你应该有所准备。

恐怖袭击的发生也是同样的道理。洛克比空难以及俄克拉荷马城爆炸案就相当于 7 级大地震，它们本身的破坏力已经足够大了，却还预示着有可能发生破坏力更大的袭击，比如“9·11”恐怖袭击，而这样的袭击则相当于 8 级地震。所以，“9·11”恐怖袭击事件并不是一个离群值，相反，它是涵盖更广的数学模型的一部分。

用统计学的方法测量恐怖主义

在用统计学方法研究恐怖主义之前，我们需要明确恐怖主义的定义。这个过程稍显复杂。弗拉基米尔·列宁曾经说过：“恐怖主义的目的是为了使人恐惧。”这个观点看似普通，其实内涵深刻：恐怖主义并不是单纯地想让死亡人数最大化；相反，恐怖分子是想使某个群体的恐惧情绪最大化，从而让该群体转变行为方式。死亡和破坏都是达到这一目标的手段。拉姆斯菲尔德也曾对我说过：“恐怖分子可能会通过杀人达到目的，但杀人并不是最终目的。”

全世界的暴力行为各式各样，学术专家致力于为恐怖主义下一个明确的定义，以区别于其他暴力威胁行为。某个热门恐怖事件数据库采用了如下定义：恐怖主义是有目的、有组织的、实际发生的、极具威胁的恐怖行为，必须由“低于国家级别的行动者”执行（即不能由某个主权政府本身执行）。另外，这类行为都包含一定的恐吓或胁迫的成分，为的是引起事件关注者的恐惧，而不仅针对直接受害者。恐怖主义旨在实现

某种政治、经济、社会或者宗教目的。

最符合上述标准的恐怖主义也最为今天的我们所熟悉，起源的时间在现代。加利福尼亚大学洛杉矶分校的政治科学家戴维·C·拉波波特将这类恐怖主义的起源时间定位在 1979 年，即伊朗革命发生的那一年，他还将恐怖主义与宗教极端主义相联系。针对西方国家及西方利益集团的恐怖袭击浪潮来势汹涌，数量激增，1979~2000 年，针对北约国家的恐怖袭击共增加了 3 倍之多。

然而，大多数此类恐怖事件即便引发了死亡，数量也不大。从 1979 年伊朗革命到 2001 年“9·11”恐怖袭击事件，针对北约国家的恐怖袭击共计发生了 4 000 多次（包括成功的和未遂的）。而这期间，一半的死亡人数仅由 7 次恐怖袭击造成。印度航空公司爆炸案、洛克比空难和俄克拉荷马城爆炸案这 3 次最严重的袭击造成的死亡人数就占了死亡总数的 40%。

有些事件发生的频率低，但影响深远，这一模型体现了幂律分布的特征，地震就遵循了这一规律。克劳塞特认为恐怖袭击同样如此。

我们画图表示恐怖袭击发生频率与死亡人数的关系，如图 13–4 所示，纵轴表示恐怖袭击发生的频率，横轴表示死亡人数。起初，这张图似乎作用不大。但从图中你能清楚观察到幂律分布：袭击次数随着发生频率降低而迅速减少。这是一条非常陡峭的斜线，似乎掩藏着什么信号：小级别的袭击发生频率高，大级别的袭击数量发生频率低，图中似乎都没有什么剩余空间留给中等级别的袭击了。于是“9·11”恐怖袭击事件看上去是个离群值。

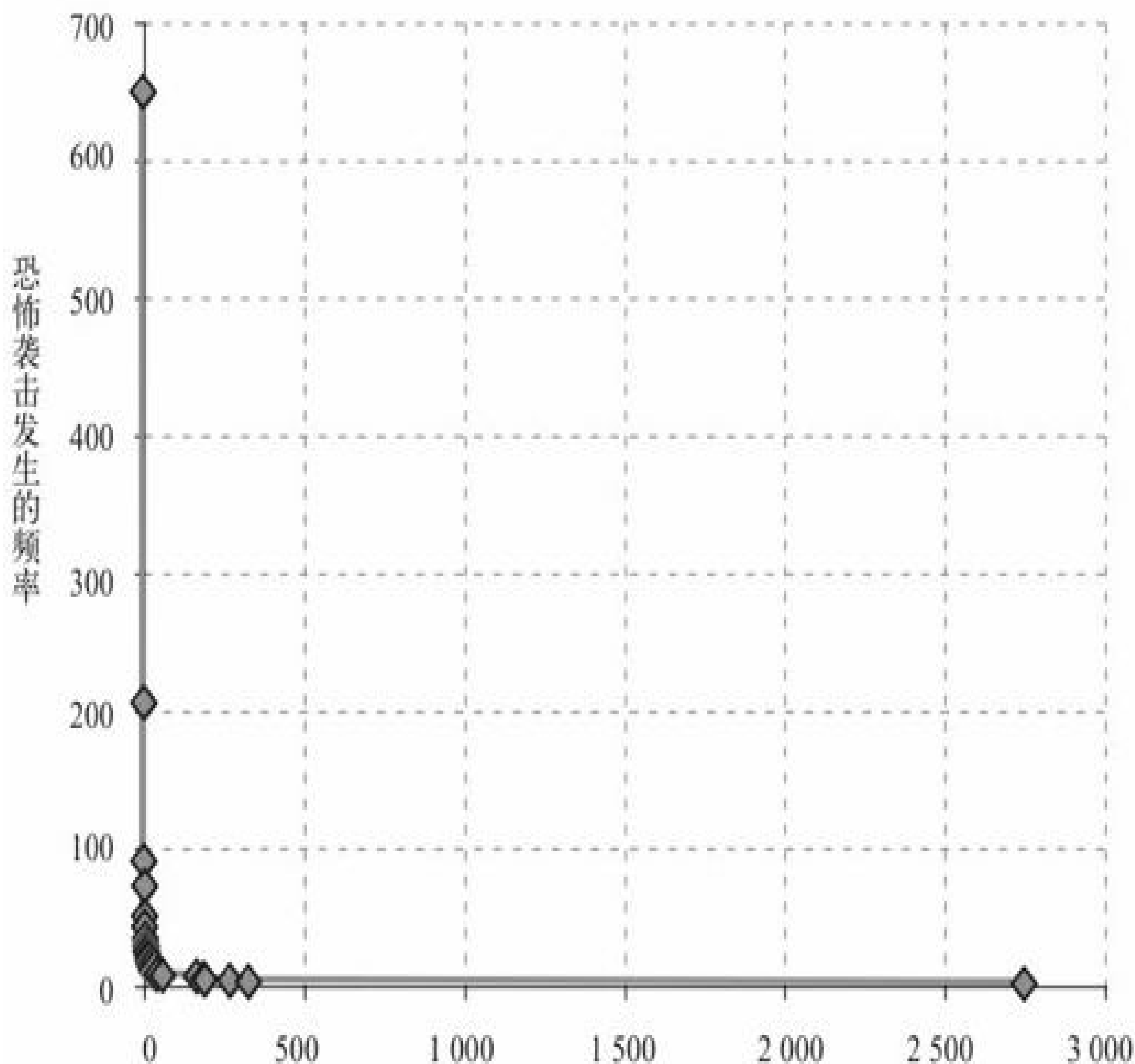


图 **13-4 1979-2009** 年北约国家发生的恐怖袭击的频率与死亡人数关系图（线性标尺）

然而，与描绘地震的情况一样，当我们用对数标尺来画图时，这些数据就变得更容易理解（更具体地说，如图 13–5 所示，用双对数标尺，即纵轴和横轴比例尺也都是对数关系）。需要强调的是，我除了让这些数据看上去更清晰之外，并没有篡改数据，它们所包含的信息是相同的。但是，原本杂乱随机的数据现在看起来整齐多了。用双对数标尺画图时，恐怖袭击发生的频率和危害程度间的关系几乎成了一条直线。事实上，这就是典型的幂律分布：当你用双对数标尺描绘符合幂律分布的事物时，得到的图像基本上呈一条直线。

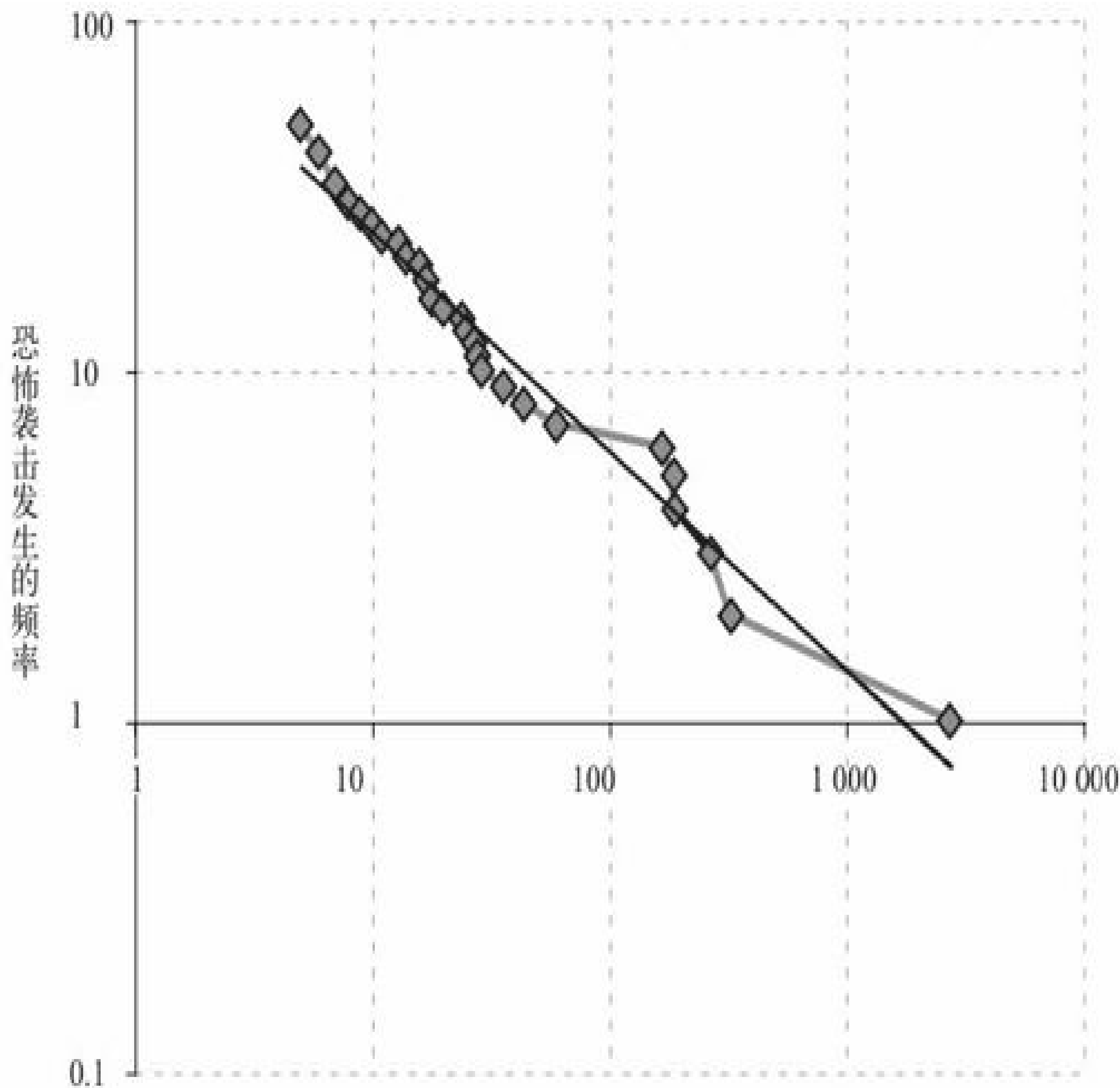


图 **13-5 1979-2009** 年北约国家发生的恐怖袭击的频率与死亡人数关系图（对数标尺）

在预测未来风险的等级时，幂律法则具有一些重要特性。其中很特别的一点是：该法则表明，比近来发生的灾难更严重的灾难是完全有可能发生的，尽管发生频率很低。例如，恐怖主义幂律法则预测某个北约成员国（不一定是美国）在 1979~2009 年这 31 年间会经历约 6 次恐怖袭击，死亡人数至少达到 100 人。（这个估值很接近实际数据，这一时期实际发生了 7 次恐怖袭击。）同样，该法则还表明每 22 年会发生一次伤亡人数达到 1 000 人等级的恐怖袭击。而造成大约 3 000 人死亡的“9·11”级别的恐怖袭击大约每 40 年会发生一次。

然而，用统计学描述历史算不得什么成就。当然，在“9·11”恐怖袭击事件发生之后，统计模型也会据此做出调整，这是有可能的。但克劳

塞特的方法曾表明，提前预测恐怖袭击或许是有可能的，这到底能否实现呢？

“9·11”恐怖袭击事件在一定程度上扭转了我们先前的观念，即认为这样严重的恐怖袭击发生的概率极低。“9·11”恐怖袭击事件如同近年来发生的一系列大地震一样，向我们表明高级别的风险威胁要比预想的更普遍。但是，早在“9·11”恐怖袭击事件发生前，幂律分布就得出了结论，像“9·11”级别的恐怖袭击显然是很有可能发生的。如果在“9·11”恐怖袭击事件之前的数据——从 1979 年伊朗革命到 2001 年 9 月 10 日所有现代恐怖袭击的数据——收集过程中运用幂律法则，就会得出这样的预测：“9·11”级别的恐怖袭击在北约国家中每 80 年会发生一次，我们一生大约会遇到一次。

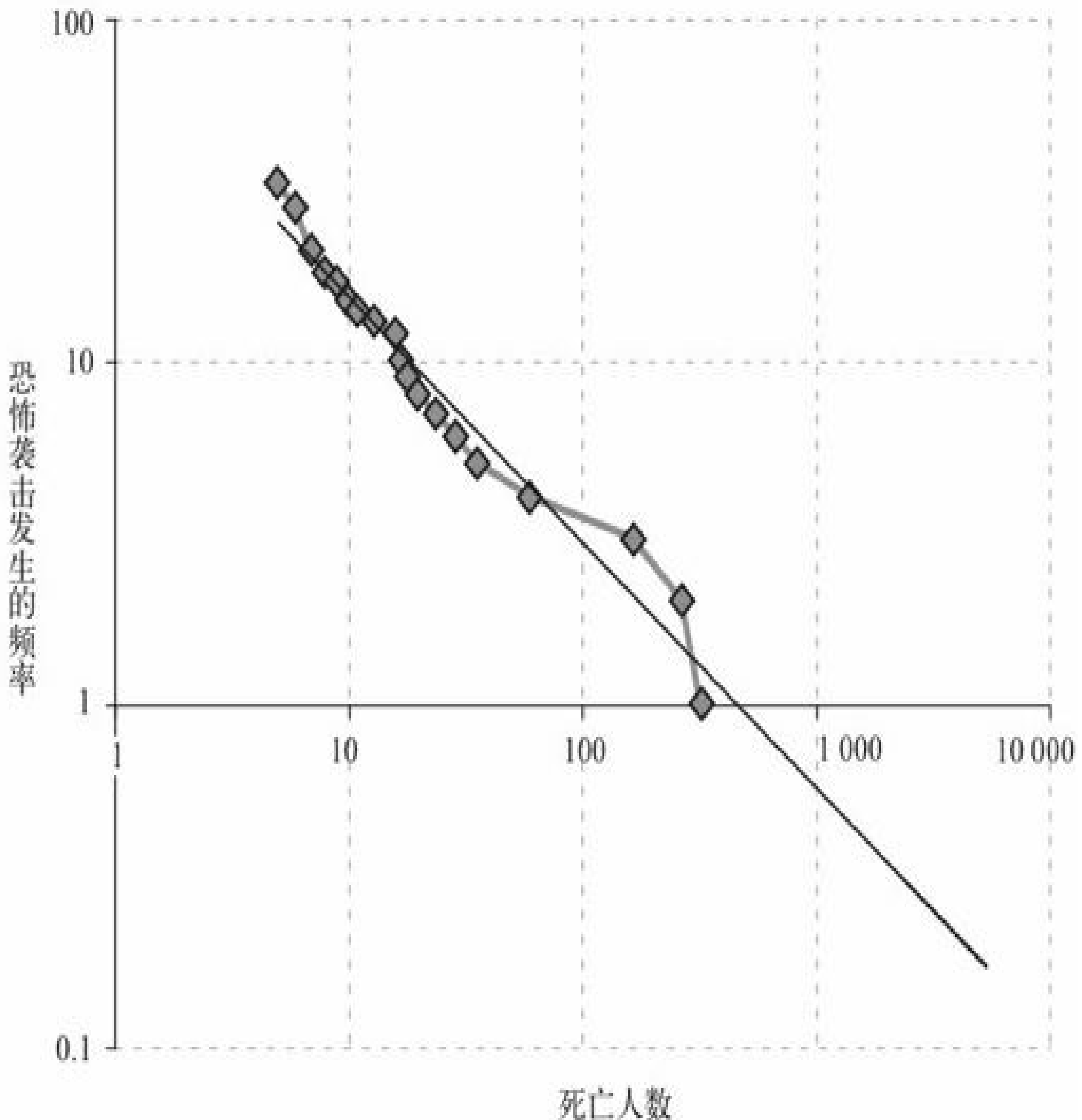


图 **13-6 1979** 年 **1** 月 **1** 日**-2001** 年 **9** 月 **10** 日北约国家发生的恐怖袭击的频率与死亡人数关系图

这个方法并不能告诉我们任何具体信息，比如具体在何处、何时将会发生恐怖袭击。它显示的是一个长期趋势，就像加利福尼亚地区地震发生频率的走势一样。但是与地震不同，恐怖袭击是能被阻止的，这对克劳塞特的假设是一个很重要的限制条件。

这些数据确实表明“9·11”级别的恐怖袭击不应该是想象之外的事件。幂律法则显示，将来完全有可能发生比已发生的恐怖袭击更严重的袭击。而我们对这些事件的无知误导了我们对其发生概率的判断。

9 级恐怖袭击很可能会发生

如果“9·11”恐怖袭击事件相当于 8 级大地震。那更严重的袭击是否相当于 9 级大地震呢？克拉塞特的方法让我们有理由相信更严重的恐怖袭击是有可能发生的，将导致数万人甚至数十万人的死亡。这种袭击令人厌恶，但其方式却很容易让人识别，很有可能包含破坏力极强的武器，尤其是核武器。

幸运的是，全球还未发生过大型核战争。1945 年，美国投放在日本广岛和长崎的原子弹造成 20 万人丧生。有人估算，若在纽约主要港口引爆一枚同等威力的原子弹，大约会夺去 25 万人的生命。自 1945 年后，核技术一直在改进。若用更加大型、更加现代的核武器袭击曼哈顿市中心，将导致约 100 万纽约市民丧命，并摧毁价值上千亿美元的房产。若同时对纽约、华盛顿、芝加哥和洛杉矶发动恐怖袭击，会杀死约 400 万美国人，这与奥萨马·本·拉登的既定目标是一致的。

这些估算大体上反映出最严重的灾难的后果，比“9·11”恐怖袭击事件严重百倍，其发生的概率已成为美国国家安全部门激烈争论的焦点。

关于这类灾难，来自哈佛大学的政治科学教授格雷厄姆·阿利森，给出了一个较为悲观的预测。阿利森曾服务于美国总统里根和克林顿的行政团队，他所著的有关古巴导弹危机的著作和论文被其他学者引用了上千次。所以，当阿利森想要表达观点时，他的同行总会洗耳恭听。

2004 年，阿利森得出一个惊人又悲观的预测：“未来 10 年美国会受到核武器的威胁，这一点不是没有可能，而是很有可能，”阿利森在书中如是说道。阿利森是这样描述他的预测的，他假设我们仍处于“现在这个世界”，存在一些危险的恐怖组织，各地都存在劣质的核材料，而美国政策决策者却忽略了这一问题。

如今，“9·11”恐怖袭击事件已经过去 10 多年了。但 2010 年我和阿利森交谈时，他再次重申他觉察到的恐怖袭击的严重程度。事实上，阿利森很认真地在思考这个问题。在与时代广场一个街区之隔的《纽约时报》报社，我在自己的办公室给阿利森打了一个电话，他告诉我说，当他身处时代广场时，总会有点小紧张，他不确定自己是否愿意每天去时代广场附近上班。

阿利森关于概率的评估，并不是来自某个统计模型。相反，这一结论的“依据正是他下赌注时的基本准则”。为什么阿利森觉得风险这么高？“这个过程就像仿版的老夏洛克·福尔摩斯对犯罪动机、手段和机会的分析一样。”阿利森如是说。

对阿利森而言，恐怖分子的动机是很容易识别的。奥萨马·本·拉登曾

说过，他想要谋杀 400 万名美国人，并且只有借助核武器才能实现这一目标。阿利森称“基地”组织的运转为“盛大的表演”，这些表演就是那些偶然发生却骇人听闻的袭击，会夺去数百名无辜平民的生命。美国中央情报局已经捕捉到信号，表明“基地”组织在“9·11”恐怖袭击事件之前就曾谈论过“美国版广岛”。

阿利森所说的机会，是指恐怖组织向美国境内走私武器的能力。阿利森认为恐怖组织肯定干过武器走私的勾当。阿利森问我：“犯罪分子每天是如何进入美国各大城市的？”美国有超过 3 700 个港口，每年接纳的货物集装箱有 600 多万个，但海关工作人员实际检查过的集装箱只占其中的 2%。“说来也许你都不信，这些人总能把武器藏在一捆大麻中蒙混过关。”阿利森半开玩笑地说。

于是，阿利森主要关注恐怖分子的手段，即恐怖组织获得核武器的能力。如果我们想降低核武器版“9·11”恐怖袭击事件发生的概率，就得对恐怖分子的手段实施有效控制。

许多专家认为，目前全世界约有 20 000 个核弹头，比 19 世纪 80 年代顶峰时期的 65 000 个少了很多。从理论上讲，核武器的威胁来自目前拥有核武器的 9 个国家——甚至美国历史上也曾发生过 11 件核武器追踪记录丢失的案例——其他国家可能正在发展核武器。但是，阿利森的忧虑主要源于两个拥有核武器的国家：俄罗斯和巴基斯坦。

阿利森认为，来自俄罗斯的威胁已有所降低，部分原因在于一些计划的成功实现，比如参议员萨姆·纳恩和理查德·鲁格发起的项目有效抑制了俄罗斯的核武器数量。俄罗斯本国的核武器数量从 1985 年顶峰时的 30 000 件减少到目前的 11 000 件。

俄罗斯的威胁已经减弱，但是巴基斯坦的威胁却增强了，甚至是急剧增强。“如果你绘制一幅有关大规模杀伤性武器和恐怖主义的分布地图，那所有路径都会贯穿巴基斯坦。”阿利森描述道。

当然也有人反对阿利森的观点，迈克尔·李维就是其中之一，他在纽约市外交关系协会工作，我曾去他的办公室采访过他。像亚伦·克劳塞特一样，李维也有着古怪的背景：在普林斯顿大学研究过理论物理学，曾担任深受美国观众喜爱的电视剧《24 小时》的技术顾问，该剧描绘了一群恐怖分子试图在洛杉矶使用核武器的故事。

李维认为恐怖袭击确实有可能发生。“初到此地时，我做的第一件事就是在地图中圈出中央车站，看看如果一包 1 000 吨的炸药在中央车站被引爆，是否会影响到我的公寓。”但是，李维认为阿利森的预测结果太过悲观，对其假设中的许多前提条件也提出过质疑。

首先，李维认为阿利森关于恐怖分子动机的想法太过想当然。李维

的意思不是否认“基地”组织有炸毁曼哈顿的企图。其实，无论是恐怖组织还是个别恐怖分子，他们都有各种各样的动机，只是根本没有努力去实施这些动机，因为他们怀疑自己没有足够的执行能力。在李维眼中，恐怖组织非常重视其阴谋得逞的概率。一次失败的尝试可能会暴露整个组织，引来美国及其他国家更密切的侦察；一次失败的袭击还会让该组织失信于现有成员和潜在的新晋成员。恐怖组织的根基是非常脆弱的，其状况就如同一家新开的餐厅一样，90%的新餐厅在开业第一年都折了本。另外，恐怖组织的招募信息在很大程度上是通过说服其成员相信该组织能让他们得到回报。尤其是“基地”组织，它发动的各类袭击成功率都很高，包括“9·11”恐怖袭击事件，也许这便解释了其长期存在的原因。可当某个恐怖组织实现回报的能力遭到质疑时，组织成员可能就会另投新主。

核打击很复杂，也很难实现。但这不代表恐怖组织骨子里反对复杂的计划，“9·11”恐怖袭击事件就足足准备了 5 年。但是，计划越复杂，所需的合作就越多，参与者也更多，他们每人都背负着巨大的风险，可能会被反恐部门侦察到，可能会叛变组织。另外，核武器袭击还需要许多高度专业化的技术知识，远比“9·11”恐怖袭击事件中那 4 个恐怖分子学开波音 767 飞机复杂得多。核物理学家数量本来就不多，那些能被恐怖组织信任的科学家更是少之又少。“如果他们发现某人拥有本科工程学位，就会让他负责一个核研究团队的工作，但我认为这种做法未见得能有什么结果。”李维说道。

最后，恐怖袭击的目标不一定就是杀更多的人，而是通过震慑手段令人改变其行为。核打击的确会夺走更多人的生命，十分可怕，但不一定就比“9·11”恐怖袭击事件可怕百倍、千倍。核打击这种手段成功概率低，这样的手段不会成为恐怖分子的主要选择。

而美国国家安全部门的其他人像拉姆斯菲尔德一样，更担心生物袭击。生物袭击对专业知识和技能的要求比核武器略低，造成的恐惧情绪却会无声地蔓延开来。人们较少遭受这类袭击。这些生物病毒具有传染性，比如重新传播天花病毒，这会使人们的恐惧心理持续几周或几个月，学校商店纷纷关门，医院实施隔离，出入境戒严。“9·11”恐怖袭击事件发生几天后，纽约市就逐渐恢复了秩序，而生物袭击过后基本上不会这么快就重新步入正轨。

拉姆斯菲尔德说：“生物武器与核武器不同。当然这种武器也会令我们敬而远之。生物武器具有传染性，还会代代相传，改变基因。人们对生物武器的恐惧与他们对核武器和化学武器的恐惧完全不同。”

生物武器造成的死亡人数难以预测，就像所有传染病的传播情况在

其爆发之前很难预测一样，而最糟的情况无疑是非常可怕的。2001 年一次名为“黑暗冬季”的模拟演习设想有 300 万美国人感染天花病毒，其中 100 万人可能死亡，恐怖分子只要同时在俄克拉荷马城、费城和亚特兰大的购物中心投放病毒，就能成功造成上述恐慌。

克劳塞特认为 9 级恐怖袭击可能会发生，但袭击的方式是不可知的，只是提出了一种可能的方式。例如，根据 1979~2009 年期间所有袭击造成的死亡人数来判断，依据类似克劳塞特的预测模型，就能推断出未来 10 年北约国家遭遇至少造成一万人死亡的恐怖袭击的概率约为 10%，造成 10 万人死亡的概率为 3%，造成 100 多万或更多人死亡的概率为 0.6%。

我们需要对这些预测结果进行慎重处理。现实中存在大量的不确定性，对极端严重的恐怖袭击发生的概率进行预测时尤其如此，技术版本不同，产生的结果也不尽相同。

恐怖主义和地震之间还可以进行另外一项有益的对比。古登堡–里克特定律认为，从长期来看，地震级别每增加一级，其发生频率就会降低 10 倍，而释放的能量则呈指数增长，约增加 32 倍。6 级地震释放的能量是 5 级地震的 32 倍，7 级地震释放的能量是 5 级地震的近 1 000 倍。

由此可见，地震级别增加一级，其释放能量的增速要比其发生频率高很多。假设发生 10 次震级为 6 级的地震才会发生一次 7 级地震，而这一次 7 级地震的危害就相当于 10 次 6 级地震的总危害。确实，低发的地震释放的地质能量会引发板块的一次大运动。1906~2005 年发生了 3 场大地震——1960 年智力大地震、1964 年阿拉斯加大地震和 2004 年苏门答腊岛大地震——它们的威力相当于这 100 年间全世界所有地震所释放的能量总和的一半。所以，地质学家和应急预案规划者主要担忧大地震的发生。当然，小级别地震若发生在错误的时间和错误的地方，也会引起巨大损失（如 2010 年海地 7 级大地震），但大级别地震仍是我们必须担忧的对象，尽管它们发生的频率很低。

再来看恐怖袭击。仅“9·11”恐怖袭击事件就夺去了 2 977 条生命——这还不包括恐怖分子的死亡人数在内——这比 30 年内（1979~2009 年）其他北约国家死于恐怖袭击的总人数还多（图 13–7）。而一次核打击或是生物袭击引发的死亡数，又会大大超过“9·11”恐怖袭击事件。

所以说，即使李维的观点是对的，即这些袭击发生的概率极低，但它们的危害却是无穷的。幂律法则估测造成百万人死亡的恐怖袭击——比如在时代广场引爆一枚原子弹——发生的年均概率只有 1/1 600。但这个比例可以这么看，每 1 600 年会死 100 万人，即每年死亡 625 人，这个数字远远大于北约国家自 1979 年以来平均每年死于恐怖袭击的人

数——180 人。提及恐怖主义，我们既要大胆评估严重袭击发生的概率，也要大胆构想如何降低风险。因此，那些指向严重袭击的信号应该被放到更重要的战略位置上。

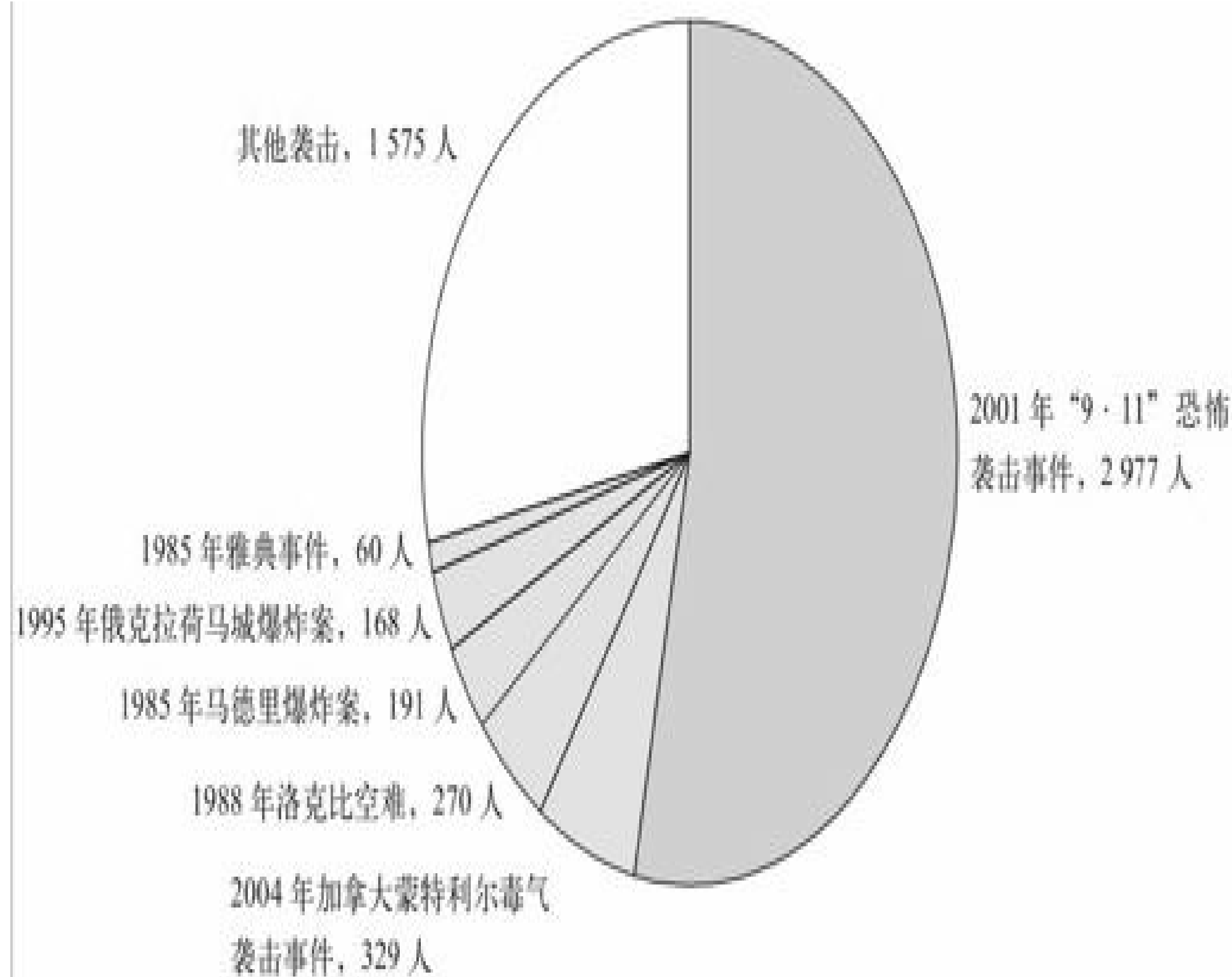


图 **13-7 1979-2009** 年北约各国恐怖袭击造成的死亡人数

关注大级别袭击的数学论据，在一定程度上减轻了国家安全工作者的日常工作。1982 年，社会科学家詹姆斯·Q·威尔逊和乔治·L·凯林引入了遏制犯罪的破窗理论，即通过集中处理轻度犯罪，比如破坏公物或轻度贩卖毒品罪，警方能有效树立法制风气，从而遏制更严重的犯罪。证明该理论的优点的论据相当复杂。

破窗理论受到美国警方的热烈欢迎，从洛杉矶到纽约的警察都接受了这一理论，因为该理论降低了警方的工作难度，并且为他们提供了更容易达成的目标。抓捕一个 16 岁的少年吸毒犯要比破获一起汽车盗窃案或是阻止一场谋杀容易得多。人人都喜欢居住在更干净、更安全的社区。但是，破窗理论到底是否只是粉饰门面，这一点还不明确。

与破窗理论类似的是被公共安全专家布鲁斯·施奈尔称为“安检表

演”的安全措施，后者看上去很厉害，但实际的反恐效果并不大。商用航线所要求的复杂琐碎的安全措施正属于此类情况。但这绝不意味着我们就完全不需要担忧机场的安全了，机场一直都是众多恐怖袭击的目标，而且恐怖主义也不乏一些盲目的模仿者——有些撞机事件和恐怖主义完全无关。在 21 世纪最初的 10 年，每 2 500 万名乘客中只有一名死于美国商业航班上。就算每年飞行 20 次，死亡的概率也只比被闪电劈到的概率大一倍而已。

为什么恐怖分子不去炸购物中心？

前文中提到的那些琐碎的安全措施多半是为了阻止愚蠢的恐怖分子，当然确实存在少数这样的恐怖分子。一个聪明的恐怖分子或许能凭借智慧逃过这些措施，或许他会将注意力转移到更脆弱的袭击目标上，比如公交车或火车。事实上，他只要把目标锁定在检票口就行，那儿人很多，安保却不到位。恐怖分子早就注意到这一点了：2011 年，莫斯科多莫杰多沃机场国际航班到港旅客区发生了一场自杀式爆炸袭击，造成 35 人死亡。

其实，除了交通系统之外，安保措施薄弱的地方还有很多。但是为什么恐怖分子不去炸购物中心呢？

之所以没有发生那么多恐怖袭击，其中一个原因也许是没有那么多恐怖分子。要清点恐怖分子的人数是相当困难的，但是有一个常被援引的估值：“基地”组织人数最多时也只有 500~1 000 人，其中包括一些随从和狂热追随者，还包括参与组织各项非暴力功能运转的成员，例如在“基地”组织中，一些胆小的家伙就得负责在该组织的网络瘫痪时对其进行重新修复。来自卡内基梅隆大学的凯思林·卡莉以叛乱组织的社交网络为研究对象。她告诉我，极端分子里只有 1%的人真的很极端。确实，去做本·拉登的信息技术顾问促进“全球圣战运动”，要比当人体炸弹冲入人群炸毁自己容易得多。

然而，在提出这类问题时，我们还是得十分认真，因为我们可能又会将不熟悉的事误认为是不可能发生的事。“恐怖分子为什么不去炸毁购物中心”这个疑问在以色列人看来荒谬无比，因为那里经常发生购物中心被袭事件。

众多批评指向克劳塞特的幂律法则，其中之一便是：与地震不同，恐怖主义是可以通过人为干预被制止的。

克劳塞特的研究表明，幂律法则并非与恐怖主义势力和反恐力量之间的较量没有关联，而是很可能因其而起。也许恐怖主义和社会之间存在着某种权衡，一种自由和安全之间的平衡，虽然这种平衡状态会随着

时空的变化而变化，但它确实是存在的。如果想生活在一个自由的社会，不管我们是否愿意承认，都必须接受一定的恐怖主义风险。

拉姆斯菲尔德告诉我："恐怖主义此消彼长。或许我们是最容易受恐怖主义攻击的群体。我们都是自由的人，因为这就是我们的本性。我们期望每天都能够过自由的日子，早晨起床出门送孩子上学时不用四下查看，唯恐被杀或被炸。但如果你总是提心吊胆，一改自由公民的行为举止，那就正中恐怖分子的下怀了。"

尽管以色列比美国更容易成为恐怖分子的袭击目标，但是以色列人并没有活在对恐怖主义的恐惧之中。2012 年一份对以色列犹太人的调查表明，只有 16%的人认为恐怖主义是最可怕的事，而更多的人称他们对以色列的教育体系最为担忧。

没有一个以色列政治家会直接宣称自己能够容忍小规模的恐怖袭击，但是这个国家的确能够容忍这一点。之所以容忍是因为该国的特殊国情——这里的每个人对恐惧都已经麻木了，这正合恐怖分子之意。以色列国家安全战略中的关键元素，就是如何在袭击之后恢复人们的生活秩序。比如，警察通常都会在 4 个小时之内清理现场，让每个人继续做自己的事。小规模的恐怖袭击在以色列被视为普通犯罪，不算恐怖威胁。

以色列真正不能容忍的是潜在的大级别恐怖袭击（比如，邻国对其进行大规模杀伤性武器袭击）。有证据表明，以色列的反恐方法卓有成效：以色列是唯一能让克劳塞特的曲线变弯的国家。如果用幂律分布的方法绘制出以色列死于恐怖袭击的人数（图 13–8），你会发现，实际发生的大规模恐怖袭击的次数明显比幂律法则预测的少，自 1979 年以来就没有发生过造成超过 200 人死亡的恐怖袭击。以色列恐怖袭击幂律分布图很好地证明了它在国家安全战略上的选择是卓有成效的。

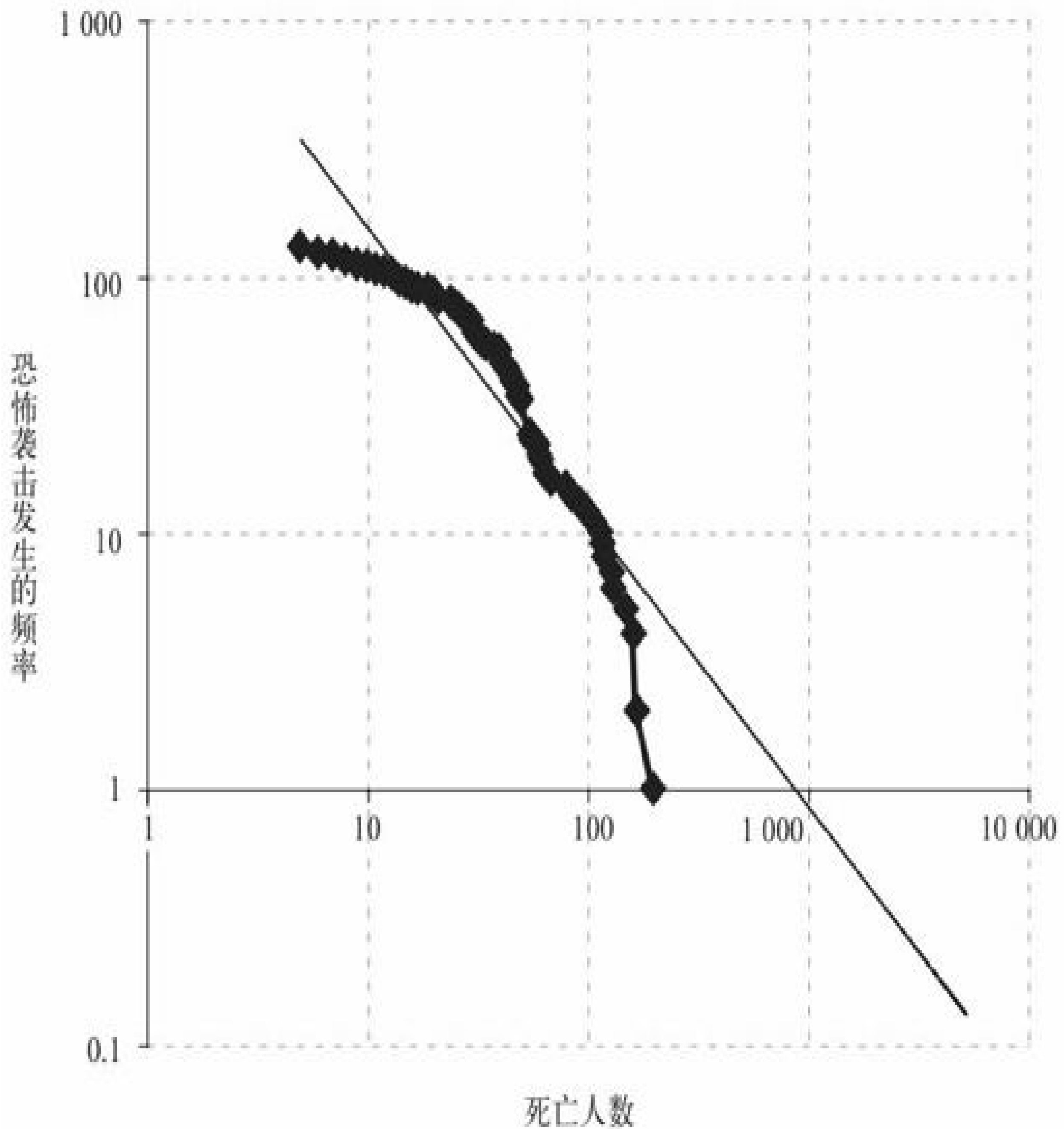


图 13-8 1979-2009 年以色列发生的恐怖袭击的频率与死亡人数

如何辨识恐怖袭击的信号？

不管我们的战略选择是什么，也不管在安全和自由之间如何做权衡，我们都要从寻找信号开始。好的情报机构是反恐的第一道防线。

美国发动伊拉克战争的既定目标之一，是阻止伊拉克继续发展大规模杀伤性武器。当然，那里实际上并没有大规模杀伤性武器。关于进军

伊拉克这一决议的几份独立分析认为，情报机构并非迫于小布什政府的压力才提供了错误信息，而是当时包括中央情报局在内的这类机构都达成共识，认为伊拉克正在积极开展大规模杀伤性武器计划，而且这些机构以几种重要的方式将这些情报误传给美国民众。

尽管这种观点有可取之处，但小布什政府告诉了公众什么、他们自己相信了什么、他们从情报官员那里知道了什么，我不确定这三者间是否能有这么大的区别。同其他类型的预测一样，在该领域的信号分析中，从错综复杂的数据中很容易看到你想要的数据。而当我们过于想要发动一场战争时，不可靠的信号来源也会被解读成可靠的，例如伊拉克战争发生前，伊拉克工程师拉菲德·阿尔贾纳比——代号“曲线球”——为了推翻伊拉克独裁者萨达姆·侯赛因，谎称自己过去服务的种子工厂是一座制造生化武器的秘密工厂，随后他承认自己编造了这一个虚假信息。

在本书提到的国际象棋中，人类往往只关注一种或两种可能的走法而置其他更好的走法于不顾。而电脑就很少犯这种错误，因为它能对所有可能的走法进行检验，并能在几秒钟内识别出诸如鲍比·菲舍尔、加里·卡斯帕罗夫等高手用过的绝招。但电脑对每步棋的评估也不会都同样深入，它的权衡方式与人类有所不同。电脑善于广撒网，而不是寻找一个完美的解决方法。

若我们总期盼世界只包括可能和不可能两个组成部分，且两者间界限鲜明，那我们在预测时就会过于自信或受未知的未知限制。既然不善于预测，那我们就应该进行对冲预测，而不是如菲利普·泰洛克口中的“刺猬”一样，死守单一的假设。

也许只有国家安全领域需要验证这么多反恐技巧。正如布鲁斯·施奈尔所言，安全问题的核心就是，各类连接中最薄弱的环节定义了我们的任务。如果你家前门是用钛镀的，还配备了武装警卫和一大群公牛，但你家还有一扇侧门，无论哪个小毛贼都能轻易从这里进入，那再好的前门也只是摆设而已。这些威胁是非对称的：当日本舰队鱼贯进入美军的防御盲点，发现美军大部分军火库都处在同一位置且极易被攻击的时候，即使用上美国海军在太平洋的全部力量也于事无补。这就是为什么在“珍珠港事件”、“9·11”恐怖袭击事件上我们产生了认知分歧。而其实敌人打击的目标是可预测的，即我们最不希望被打击的地方。

本书中，我推荐的一些思维方式有益于国家安全分析。比如，完全没把握做出某一决定时，贝叶斯定理的概率式思考方法也许更合适些，该思考方法鼓励我们同时提出许多不同的假设并考虑其发生的概率，当遇到与其有关的新信息时，还需要时常更新你的预测。

在“9·11”恐怖袭击事件发生之前我们差点就识破“基地”组织的阴谋了。当时美国逮捕了扎卡里亚斯·穆萨维，这个伊斯兰极端分子对驾驶波音 747 飞机异常感兴趣。如何证明他的清白呢？或许他过去曾是一个有钱有闲的飞行爱好者。但如果我们为“恐怖分子可能劫持飞机，冲撞大楼”这个假设设定一个先验概率，即便概率很低，一旦发现相关情报，对此类恐怖袭击发生概率的评估值就应该大大提高。事实却正好相反，我们压根儿就没考虑过这一假设的存在，认为它对我们来说是一个未知的未知。正如“9·11”恐怖袭击事件委员会总结的那样，“整个系统都无法理解这一信息的潜在含义”，因此穆萨维的被捕未能帮助美国瓦解“基地”组织的袭击计划。

但这不是说我们的情报机构就一事无成。如果说“9·11”恐怖袭击事件侦察失败必须要遭到谴责，那么小布什政府和奥巴马政府则该受到赞颂，因为自“9·11”恐怖袭击事件之后再无袭击发生——在 11 年前，我肯定不会做出这样的预测。情报分析家如同棒球裁判一样，犯错时常被谴责，表现出色时却无人喝彩。我并不认为情报领域的失败预测是一个天大的失败，如同我在本书中所列的其他失败案例一样糟糕；考虑到该领域预测时面临的挑战，这样的预测已经算是比较成功了。

然而，正如“9·11”恐怖袭击事件委员会推测的那样，我们之所以无法成功预测恐怖袭击，最大的原因在于我们缺乏想象力。预测时，我们需要在好奇和怀疑之间寻求平衡。这两者是可以相互协调与融合的。我们越是渴望检验自己的假设，就越愿意承认我们关于世界的知识存在很多不确定，越愿意承认不可能做出完美的预测，越不会陷入失败的恐慌中，也会有更多的自由让思维驰骋。对自己不了解的事物作进一步的了解，我们也许就能做出更多准确的预测。

结束语

对美国职业棒球大联盟的游击手来说，有些比赛可以打好，有些比赛永远也无法打好，有些比赛则一定要奋力打好。球员们奋力拼搏的比赛是最精彩的，而且总能吸引人们的注意，但这样的比赛也会使人们低估游击手的能力。

德瑞克·基特一直是“点球成金”时代备受争议的主角。赛事转播员和球探们注意到基特的比赛似乎都十分精彩，从而得出结论，认为他是该赛季的最佳游击手。而统计怪才在分析这些比赛时，却发现了这种结论的缺陷。尽管基特是一个极好的棒球运动员，但他跃起接球的速度较慢，不得不俯冲上垒以弥补失去的时间。实际上，有分析数据显示，尽管曾 5 次赢得大联盟“金手套奖”，但基特其实只是一个表现极其一般的防守游击手。基特奋力拼搏的比赛得到的荣誉不会太多，因为这些比赛算不得惊心动魄，若是换了奥奇·史密斯这样的防御游击高手，拿下这些比赛可能就是轻而易举的事。

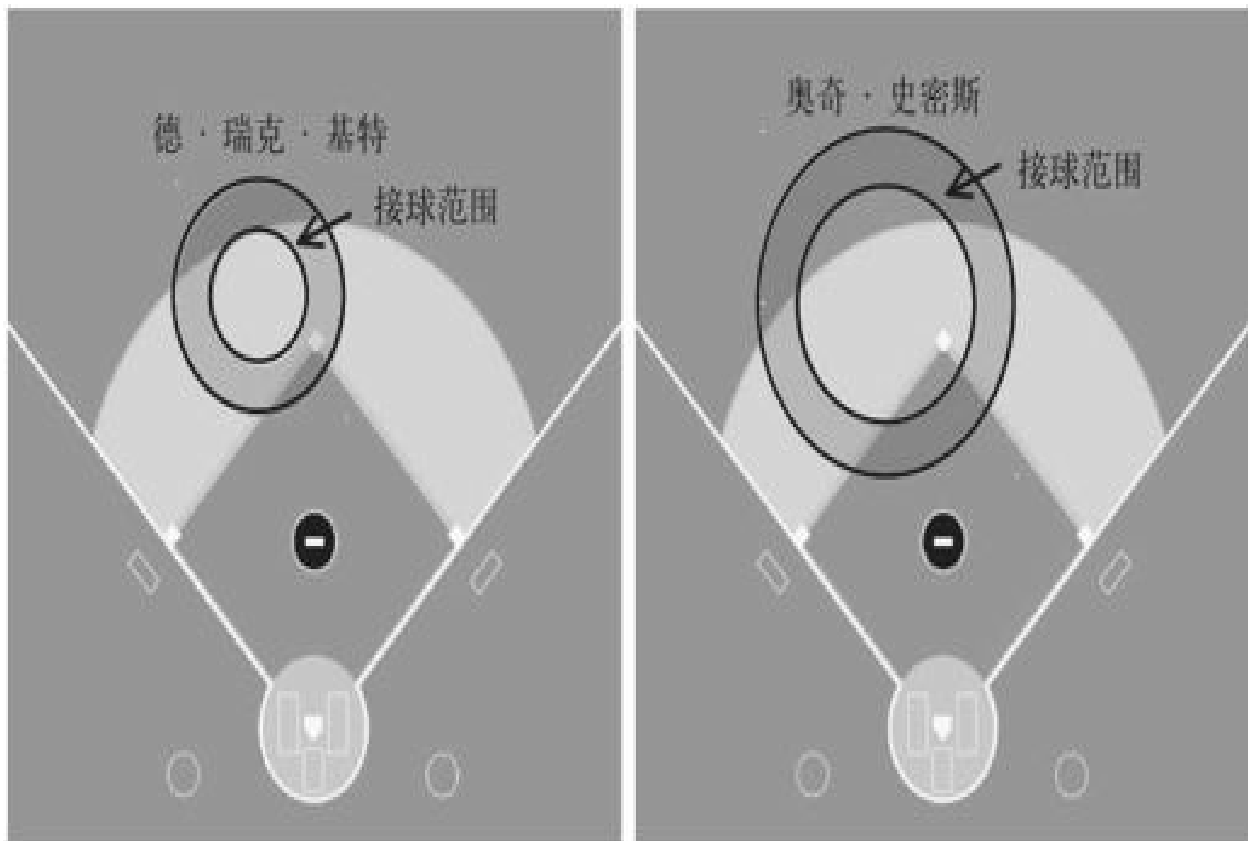


图 **1** 游击手奋力接球的范围

不论能力范围如何，总会有任务等着我们竭尽全力去完成。如果我们用最难的任务来判断自己的能力，那么轻而易举且按照常规完成的事就会被视为理所当然的了。

历史上最引人注意的正确预测来自英国天文学家埃德蒙·哈雷，他在1705 年曾经预测，一颗巨大的彗星会在 1758 年回归。曾经有很多人怀疑过他的预测，但彗星恰恰就在 1758 年回归了。在古代，彗星被视为上帝赐予的完全无法预测的事物，如今却成了有规律且可预测的事物。

天文学家预测，哈雷彗星下一次最接近地球的时间是在 2061 年 7 月 28 日，到那个时候，现在困扰着我们的自然界的许多预测难题，都会在我们的知识范围内了。

自然法则并没有改变多少，只要人类知识不断积累（自从古腾堡的印刷机问世以来确实是这样），即使不能完全领悟大自然的奥秘，我们对大自然中信号的理解也会逐渐加深。

然而，如果说科学与技术是本书的主角，那么在大数据时代，人们对科技所能完成的事恐怕过于乐观了。

没有理由认为人类活动越来越可预测，同样也没有理由认为人类活动越来越不可预测。科学使得社会变得明朗，但科学同样也使得社会组织变得更加复杂。技术完全改变了人们彼此之间的联系。1990 年发明万维网的蒂姆·伯纳斯·李对我说："正因为有了互联网，整个环境、所有方程式、所有信息的动态传播都发生了变化。"

大量的信息成倍增加，但有用的信息却非常有限，信号的比例正在缩小，我们需要找到更好的方法对信号和噪声进行区分。

本书讨论我们已知的部分较少，讨论更多的是已知与未知之间的差异，并向人们推荐了缩小鸿沟的策略——人们既需要迈出重大转变的步子，同时也得走好细微谨慎的步子。而重大的一步就是要求大家采用贝叶斯关于预测的概率式思考方法。

以概率的方法思考问题

贝叶斯定理的开始和结束，都是以对真实世界可能性的盖然论表达出来的。贝叶斯定理并不要求每个人都相信这个世界从本质上来讲就是不确定的，这一定理提出的时候正是牛顿运动定律成为科学典范的时候。然而，贝叶斯定理却要求人们接受这样的观点，那就是人对这个世界的主观看法确实是十分接近真相的。

贝叶斯定理主张的概率式思考方法，起初可能会让人感到有些不舒服。除非是玩牌或其他概率性游戏，否则人们不会以概率的方法思考问题。例如在数学课上，人们花在几何和微积分这样的抽象学科上的时间要大大多于花在概率论和统计学上的时间。而在现实世界的各行各业里，不确定性甚至会被误认为是不自信的表现。

当你第一次对概率进行预测时，可能不会特别准确。但有两条利好

消息：第一，这些预测只是一个开始，当你得到新信息时，贝叶斯定理会指导你对你的预测进行修正。第二，有证据表明，我们可以通过学习识别各种信号改进预测。例如，军队有时就会使用这些技术训练士兵，效果相当不错。医生也会用贝叶斯定理进行医疗诊断。

与电视专家相比，效仿医生和士兵的方法可能会更好一些。

大脑在处理信息时使用的是近似法。与其说这是一种既成事实，不如说它是一种生物必要性：我们察觉到的信息远多于我们有意识进行思考的信息，我们处理信息的方式是按照规律和模式对它们进行分类。

在高度的压力下，生活的规律会慢慢被揭示出来。针对“9·11”恐怖袭击事件的幸存者作过的一项研究发现，幸存者对他们所经历的事只能回忆起几分钟的细节，但对于“更大的环境”他们几乎完全不记得了。在这样的情况下，第一直觉和第一近似解可能非常不准确，无法意识到威胁的严重性。那些在强压下被迫做出决定的人，如在战场上的人，更容易成为带领其他人脱离险境的英雄。

在日常生活中，大脑也会尽可能地简化事物并求取其近似值。利用既有经验，这些简化和求取近似值的过程会成为有用的向导，构建起我们在日常生活中应用的可操作性知识。这两种方法并不完美，而我们也经常意识不到它们的粗糙性。

仔细思考以下 7 条陈述，它们与有效市场假说和个人投资者能否击败股票市场有关。每条陈述都只是上述关系的近似值（即粗略说明），但每一条都建立在最后一条的基础上，而且越来越准确。

1.没有哪个投资者能够击败股票市场。

2.长期看来，没有哪个投资者能够击败股票市场。

3.相对于其风险水平而言，长期看来，没有哪个投资者能够击败股票市场。

4.相对于其风险水平而言，考虑到其交易成本，长期看来，没有哪个投资者能够击败股票市场。

5.相对于其风险水平而言，考虑到其交易成本，长期看来，没有哪个投资者能够击败股票市场，除非他有内部信息。

6.相对于其风险水平而言，考虑到其交易成本，长期看来，几乎没有哪个投资者能够击败股票市场，除非他有内部信息。

7.长期看来，有多少投资者能够击败股票市场是一件很难说清的事，因为数据非常嘈杂，但我们知道，相对于其风险水平而言，大多数投资者都无法击败股票市场，因为股票交易存在交易成本，也就不会产生净超额回报。这样一来，除非他有内部信息，投资指数基金或许能有不错的赢利。

第一条陈述，即“没有哪个投资者能够击败股票市场”这条无限制条件的陈述，似乎极为确定。而到了最后一条则充满了不确定性，并不适合用作“车贴”式的标语，但这却是一个对客观世界更加完整的描述。

生活中处处都有近似事物，这很正常。如果你遇到一个对股票一无所知的陌生人，你告诉他，即使是在第一条陈述所描述的无限制条件的情况下，人们也很难击败股票市场，知道这一点也比一无所知好得多。

但是，一旦我们把近似误认作现实，问题就出现了，菲尔·特罗克提出的刺猬型专家就是这样。越简单的陈述似乎越符合一般情况，越能证明更加伟大的真相或是理论。然而，特罗克发现，刺猬型专家很不擅长作预测，他们总会遗漏所有能够使生活更加真实、使预测更加准确的点点滴滴。

我们的大脑有智慧，但我们生活的宇宙更是大得令人费解。以概率的方法思考问题的优点是，我们可以借这种方法强迫自己停下来，查出数据，放慢速度，仔细思考自己想法的不足之处。随着时间的流逝，我们会发现概率法会使我们的决定更加合理。

知道自己的观点源于何处

贝叶斯定理要求我们在权衡各种迹象之前，就要指出——并且是明确地指出——这一事件发生的可能性有多大。这种预测被称作“初始观点”。

我们的初始观点从何而来呢？从理论上讲，我们希望将初始观点建立在过去的经验——最好是社会经验——的集合之上。这是市场可以扮演的有用的角色之一。市场当然不是完美无瑕的，但绝大多数时间内群体判断都要优于个体判断。市场在权衡新迹象的时候形成了一个好的起点，在你还没有在某个问题上花费太多时间的情况下，尤其如此。

当然，市场并不适用于所有情况，将一些个案挑出来作为默认情况也是很有必要的。即使是常识，也可以作为贝叶斯定理的前提条件，与容易轻信的统计模型的输出结果进行比对。（这些统计模型虽然看似可以保证数学精确性，但其给出的都是近似结果，而且非常粗糙。）信息只有在恰当的环境下才会成为知识，没有环境，我们就无法从噪声中区分信号，我们对真相的追寻也会陷入错误判断的泥潭。

贝叶斯定理不能接受的，就是你假装自己没有任何初始观点的做法。人们应该努力减少偏见，但如果你说一点儿偏见都没有，反而暗示了你有很多偏见。预先陈述自己的观点——如“我的观点正源于此”——是诚信预测的方式，由此也可以认识到，我们对事实的感知是经过主观过滤的。

在不断的试错中进步

不断犯错，不断尝试，这或许是贝叶斯定理应用起来最容易的一个原则了：进行大量的预测。你可能不会将自己的公司或是生活赌在预测上，尤其是刚起步的时候，但这是唯一能够让自己取得进步的方式。

贝叶斯定理告诉我们，任何时候获得新信息，我们都应该更新自己的预测。简单地说就是，不断犯错，不断尝试。真正“拥有”大数据的公司，比如谷歌公司，并不会在建立模型的问题上花费太多时间。它每年会进行上千次实验，并在真正的客户身上检验它的想法。

贝叶斯定理鼓励我们权衡新信息时要遵守规律。如果我们的想法确实有价值，我们就应该建立可以证伪的假设来验证它们，并且将它们应用于预测当中。大多数时候，我们意识不到数据是多么嘈杂，所以对于最新的数据我们总是强加了太多个人偏见。政治记者经常会忘记，他们所报道的民调会存在误差幅度，而金融记者总是不能很好地向公众传达大多数经济统计数据是多么不精确。制造新闻的人往往是局外人。

但是，当我们在解决某个问题时，个人情感过多或是过于专业化，当事实发生改变而我们却无法改变时，就可能会产生相反的偏见。如果某个专家属于刺猬型，那么当数据与他的世界观不一致时，他可能会因为过于骄傲而不去改变自己所作的预测。各党派支持者总是期望他们的每个想法都能印在保险杠贴纸上，在他们承认自己对事实进行了过分简化之前，会经历各种错误。

验证想法的频率越高，就能越早地避开这些问题。眺望大海，等待着灵感迸发，想法就出现了，这是电影里才会有的情节。在真实世界中，即使已经准备就绪，想法也很少会出现，“大”想法就更不用说了。更加常见的情况是，我们只能凭借微小的、渐进式的，有时甚至是偶然出现的想法取得进步。

对可预测性的认知能力

预测之所以难做与其之所以重要的原因是一样的：预测是主观事实与客观事实交汇的产物。从噪声中区分信号既需要科学知识，也需要自知之明，比如，平静地承认我们无法预测的事物、勇敢地说出我们能够预测的事物，还有就是明智地区分二者的不同。

多年来，对于我们能否预测世界的观点经历了各种兴衰成败。一个简单的衡量标准就是，学术期刊中“可预测性”和“不可预测性”这两个词出现的频率。20 世纪之初，这两个词出现的频率相当。“大萧条”和第二次世界大战使得“不可预测”一词占据上风，而等到世界从危机中走出，“可预测性”一词便强势回归，在 20 世纪 70 年代达到顶峰。而最近几年，“不可预测性”一词又卷土重来，气势高涨。

对“可预测性”的认知能力受科学思潮和人类短暂的记忆力（比如记性不好的我们常会问，最近发生什么不好的事了吗？）的影响，要多于预测技能中的任何实质性改变对它的影响。我们对自己的预测有多满意和我们的预测准确与否，这两者之间或许是负相关的关系。20 世纪 50 年代，世界仍因为战争而动荡不安，局势相当难以预测，而到了 20 世纪 70 年代，人们以为自己能够预测所有事物，但实则不能，这两个 10 年相比，20 世纪 50 年代的经济和科学生产力更胜一筹。

这些态度的转变所产生的影响，已经远远地超过了学术期刊的影响。如果根据英文小说中“可预测性”和“不可预测性”两个词的使用情况绘制一张图表，可能会与图 2 所示几乎一致。一个不可预测的灾难，即使对我们不会产生直接影响，也会动摇我们掌握自己命运的信心。

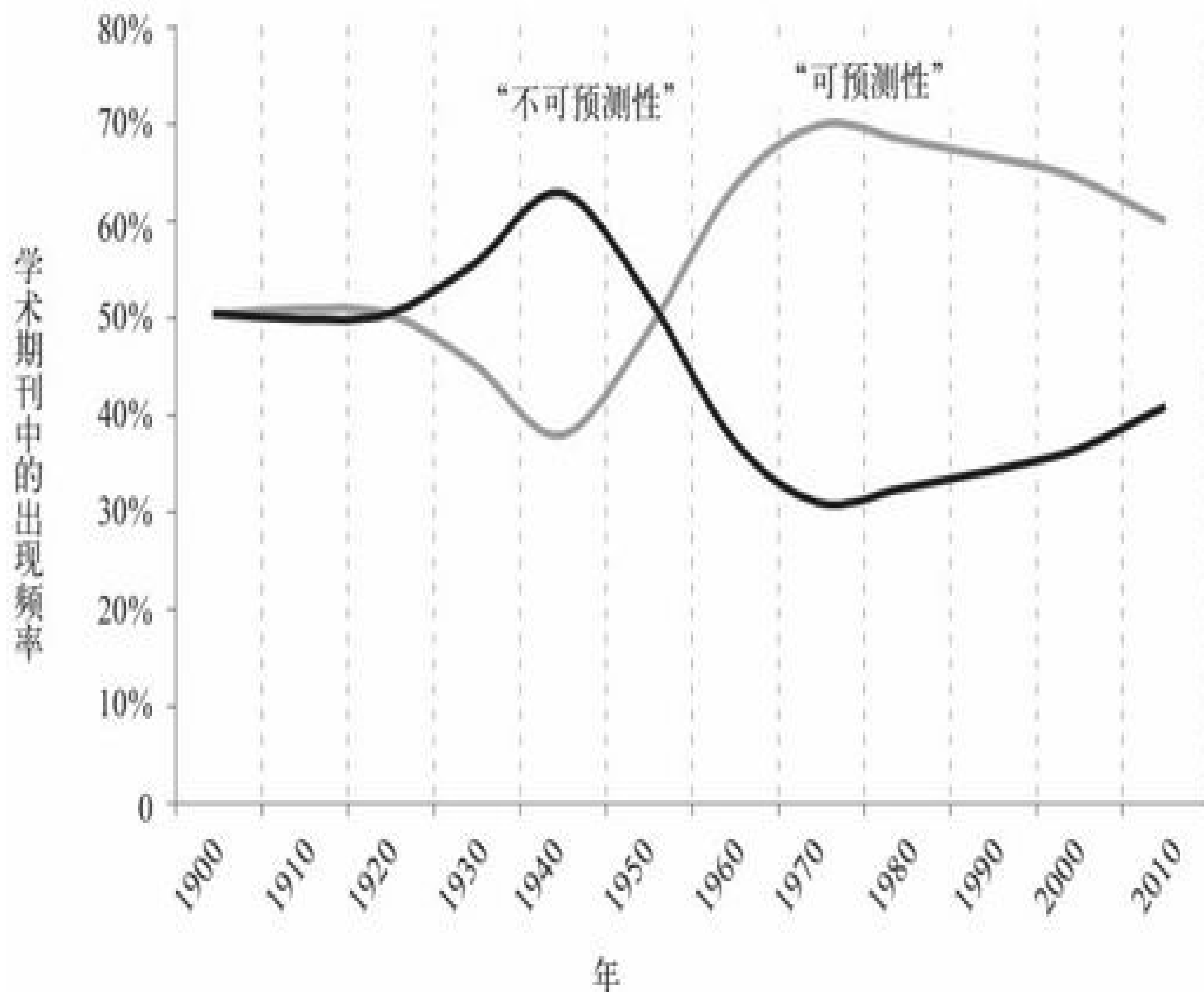


图 **2 1990-2012** 年对**“**可预测性**“**的认知能力

但我们还是倾向于认为我们所作的预测比实际的要好。新千年的前 12 年里，无法预料的灾难一个接着一个，实在是坎坷连连。我们想在灰烬中重生，不向命运低头，对我们的预测能力就应该多保持谦虚的态度，才能尽量避免重蹈覆辙。

致谢

作家约瑟夫·艾本斯坦曾经说过，完成一本书的感觉很棒，但是创作的过程却很痛苦。确实，写书需要有组织、有条理、有耐性，还需要其他许多品质，而我正缺少这些品质，撰写博客并没有帮助我养成这些品质。

因此，我得依靠那些拥有这些品质的人，他们的智慧或多或少地影响了本书的架构。

在此，我要感谢我的父母布莱恩·戴维·希尔福和莎莉·特伦·希尔福尔，以及我的姐姐瑞贝卡，谨将本书献给他们。

感谢编辑弗吉尼亚·史密斯，她在各方面都十分优秀，感谢劳拉·斯蒂克尼、安·戈多夫以及斯科特·莫耶斯这 3 位编辑对本书前景的信心。他们十分专业，编辑过的书几乎都饱受赞誉；当我需要更多的时间写作时，他们对我提出的各类借口十分宽容，对此，我深表谢意。

感谢我的文学经纪人赛德勒·克莱默，感谢他帮助我构思和运作本书的推广计划。他的建议谦虚有度，却不过分低调，事实证明他的建议总是正确的。

感谢我的研究助理艾瑞基亚·米利肯为本书投入无比高的热情，对本书在科学和技术方面的兴趣导向产生了很大的影响。感谢茱莉亚·卡明，她的组织能力帮助我顺利通过几次创作的瓶颈期。还要感谢简·卡罗莱纳和埃伦·卡罗莱纳·波特，感谢两位能在高强度下完成本书的誊写工作。

感谢艾米丽·沃图巴、维罗妮卡·温德霍兹、凯特琳·福林、阿曼达·杜威和约翰·夏普，他们得在紧迫的日程安排下完成出版任务。日程常要求他们“当日完工”，这个“当日”指的是“当晚”，而“当晚”通常指“凌晨 5 点前”，感谢他们对这些苛刻要求的理解和支持。

感谢罗伯特·高尔丁的支持和厚爱。感谢莎肖克·帕特尔、金·巴林、布莱恩·乔伊纳、凯蒂·哈尔珀、杰森·麦克林、马雷·萨勒哈、杰西卡·克莱茵等人对我的包容。有时我会就本书的相关内容喋喋不休地和他们说上好几个小时，有时又会突然消失几周。

感谢来自《纽约时报》的迈卡·科恩，他对本书的帮助无法一一尽数。

感谢《纽约时报》的老板和同仁们，尤其感谢梅甘·利伯曼、吉姆·罗伯特、戴维·莱昂哈特、莉莎·托奇、格里·马拉尼、里克·伯克、迪克·斯蒂文森、德雷克·威利斯、马特·埃里克森、格雷格·威伊斯、雨果·林格伦等同仁对我的信任，让我得闲安排本书出版周期内的各项事宜以及出

版后的各项事宜。感谢比尔·凯勒、格里·马佐拉蒂和吉尔·埃布拉姆森，是他们3位带我进入了《纽约时报》这个大家庭。

感谢约翰·赛德、安德鲁·格尔曼、汤姆·斯卡勒、艾德·季格尔、雷纳·塞克斯顿、布莱恩·麦凯布、哈尔·斯图尔特和肖恩·奎因对538网站所作的贡献。

感谢芝加哥大学的理查德·泰勒和阿尼尔·卡西亚普对本书中经济和金融相关章节的审阅。感谢戴维·卡尔、凯西·高尔丁和佩吉·阿什利提醒我一本书结尾部分的重要性，还要感谢威尔·雷普库教给我的写作准则，这一点对本书的出版帮助很大。

感谢加里·哈克贝、布兰登·亚当、雷夫·福尔斯特、凯文·戈德斯坦、基思·乌巴汗、马休·沃格尔、蕾切尔·豪泽、詹妮弗·布洛克、汤姆·肖克、李志远和马克·戈德斯坦，他们在本书创作过程中的一些关键时刻为我充当了联络员与协助者。

另外，还有许多人参与了本书书名的征集调查。在此要特别感谢以下朋友所提出的宝贵建议：乔纳·佩瑞提、安德里亚·哈纳、凯尔·罗斯、杰西·佩弗拉、鲁斯·韦尔特、布伦特·希尔福、理查德·希尔福、阿曼达·希尔福、洛伊尔·林德格尔、莱恩·林德格尔、祖本·杰尔韦哈、道格拉斯·杰斯特、贾斯汀·沃尔弗斯、J·斯蒂文·斯德帕、罗伯特·埃里克森、凯蒂·唐娜雷克、海伦·李、卡莎·波利特、杰弗瑞·图本、戴维·罗伯特斯、菲利克斯·撒曼、希拉里·伯克、希特勒·哈尔伯特、亚特·戈德哈默、戴维·卡洛儿、莎拉·罗宾森、马克思·萨维奇、迈克尔·奥哈尔、马尔可·特雷西、丹尼尔·戴维、E·J·格拉夫、保罗·斯塔尔、罗斯·威伦、杰弗瑞·豪赛、达纳·戈德斯坦、苏西·金敏、乔纳森·左斯罗福、艾维·真尼尔曼、詹姆斯·加尔布雷斯、格雷格·安瑞格、保罗·沃尔德曼，以及鲍勃·库特纳。

我所采访的一些人对本书内容所涉及的方向有重大影响，而我或许未能在书中充分提及这些人，在此，我真挚地感谢他们：丹尼尔·卡尼曼、瓦西克·拉吉里奇、亚历山大·桑迪·麦克唐纳、小罗杰·皮克尔、约翰·朗德尔、托马斯·乔丹、艾琳·埃科斯特兰德、菲尔·戈登、克里斯·福林斯凯、罗伯特·贝尔、蒂姆·伯纳斯–李、莉莎·兰德尔、杰伊·罗森、西蒙·杰克曼、戴安娜·劳德戴尔、杰弗瑞·萨克斯、霍华德·莱德勒、罗德尼·布鲁克斯、亨利·阿伯特，以及布鲁斯·比诺德·梅斯吉塔等。

我希望日后能回馈各位的帮助。或许，我首先会为上述各位买上一瓶好啤酒，而那些本应该出现在感谢名单中却被我粗心遗漏的人，更应该好好感谢他们一番。

纳特·西尔弗

纽约布鲁克林